Running head: TROUBLE WITH OVERCONFIDENCE

The Trouble with Overconfidence

Don Moore
Tepper/CMU
5000 Forbes Avenue
Pittsburgh, PA 15213

don.moore@alumni.carleton.edu
412-268-5968

Paul J. Healy
410 Arps Hall
1945 North High Street
Columbus, OH 43210-1172

healy.52@osu.edu
614-247-8876

Draft of December 3, 2007

Don A. Moore holds the Carnegie Bosch Faculty Development Chair at Tepper School of Business, Carnegie Mellon University. Paul J. Healy is at the Department of Economics, Ohio State University. The authors wish to acknowledge the financial support of the National Science Foundation (Grant No. 0451736), a CART grant from the Tepper School of Business at Carnegie Mellon University, and the staff and facilities of the Center for Behavioral Decision Research at CMU. We appreciate helpful comments on the paper from Max Bazerman, Peter Bryant, Katherine Burson, Uriel Haran, Mathew Hayward, Robin Hogarth, Daniel Kahneman, and Sabina Kleitman. Thanks to Zachary Burns for help with data collection. Most of this paper was written while the first author was visiting at the University of New South Wales, Australia. Some alternative approaches to analyzing the original data presented in this paper appear in Healy and Moore (2007). Please address correspondence via email to: don.moore@alumni.carleton.edu.

The Trouble with Overconfidence

Abstract:

This paper presents a reconciliation of three distinct ways in which the research literature has defined overconfidence: (1) overestimation of one's actual performance, (2) overplacement of one's performance relative to others, and (3) excessive precision in one's beliefs. Experimental evidence shows that reversals of the first two (apparent underconfidence), when they occur, tend to be on different types of tasks. On difficult tasks, people overestimate their actual performances but also mistakenly believe that they are worse than others; on easy tasks, people underestimate their actual performances but mistakenly believe they are better than others. This paper offers a straightforward theory that can explain these inconsistencies. Overprecision appears to be more persistent than either of the other two types of overconfidence, but its presence reduces the magnitude of both overestimation and overplacement.

Key words: overconfidence, underconfidence, overestimation, overplacement, overprecision

The Trouble with Overconfidence

Overconfidence can have some serious consequences. Researchers have offered overconfidence as an explanation for wars, strikes, litigation, entrepreneurial failures, and stock market bubbles (Camerer & Lovallo, 1999; Glaser & Weber, 2007; Howard, 1983; Johnson, 2004; Malmendier & Tate, 2005; Neale & Bazerman, 1985; Odean, 1999). In the words of one popular text on judgment, “No problem in judgment and decision making is more prevalent and more potentially catastrophic than overconfidence” (Plous, 1993, p. 217). Because of its generality and importance, overconfidence research has been broadly influential outside of psychology (Daniel, Hirshleifer, & Subrahmanyam, 1998; García, Sangiorgi, & Urosevic, 2007; Hoelzl & Rustichini, 2005; Santos-Pinto & Sobel, 2005; Statman, Thorley, & Vorkink, 2006). However, overconfidence has been studied in inconsistent ways. In this paper we examine the three different ways that overconfidence has been studied and attempt to reconcile the contradictory results from those studies.

The Three Faces of Overconfidence

A search of the PsycInfo database using the word “overconfidence” yields 365 hits, 72% of which are empirical papers. This literature has defined overconfidence in three distinct ways. The first definition of overconfidence is the overestimation of one’s actual ability, performance, level of control, or chance of success. In order to distinguish it from the other forms of overconfidence, we will call this overestimation. If a student who took a ten-item quiz believes that he answered five of the questions correctly when in fact he only got three correct, then he has overestimated his score. Roughly 64% of empirical papers on overconfidence examine overestimation.

The second variety of overconfidence occurs when people believe themselves to be better than others, such as when a majority of people rate themselves better than the median.[1] Roughly 5% of empirical overconfidence papers (12 of them) examine this phenomenon. Note that this research does not always use the term overconfidence—a search using the more specific term "better-than-average" produces another 146 articles (for a review of this literature, see Alicke & Govorun, 2005). Following Larrick, Burson, and Soll (2007), we will call this overplacement. If a student guessed that her score was the best in the class, when in fact half of the class scored better than she did, then she has overplaced her score relative to those of others.

The third variety of overconfidence is excessive certainty regarding the accuracy of one's beliefs, or what we will call overprecision. Roughly 31% of empirical overconfidence papers examine overprecision. Researchers examining overprecision typically ask their participants questions with numerical answers (e.g., "How long is the Nile River?") and then have participants estimate 90% confidence intervals around their answers. Results show that these confidence intervals are too narrow, suggesting that people are too sure they know the correct answer; 90% confidence intervals contain the correct answer less than 50% of the time (Alpert & Raiffa, 1969/1982; Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999; Soll & Klayman, 2004). For example, if a group of one hundred students were all at least 90% sure that they had scored below five on the ten-item quiz, but in fact 20% of them had scored over five, then their judgments would display overprecision.

Researchers routinely assume, either explicitly or implicitly, that the different types of overconfidence result from the same underlying psychological causes (e.g., Alba & Hutchinson, 2000; Barber & Odean, 2001; Belsky & Gilovich, 2000; Daniel et al., 1998; Dunning, 2005; Juslin, Winman, & Olsson, 2000; Kirchler & Maciejovsky, 2002; Malmendier & Tate, 2005;

Moore, Kurtzberg, Fox, & Bazerman, 1999; Odean, 1998; Plous, 1993; Stone, 1994; Stotz & von Nitzsch, 2005). For instance, overestimation and overplacement have frequently been treated as interchangeable manifestations of self-enhancement (Kwan, John, Kenny, Bond, & Robins, 2004). In this paper we ask whether the three varieties of overconfidence we have identified are merely differing operationalizations of the same underlying construct, and if not, how they are related to each other.

We first review the literature on overconfidence and identify empirical inconsistencies and methodological problems. As it turns out, many of these inconsistencies and problems result from treating all three varieties of overconfidence as if they were the same. Second, we present a new theory that can help us reconcile these inconsistencies. Third, we present an illustrative experiment that measures all three varieties of overconfidence concurrently and allows us to test some of the novel predictions of our new theory.

The Trouble with Overconfidence

There are three notable problems with research on overconfidence. The first is that the most popular research paradigm confounds overestimation with overprecision. The second is the prevalence of underconfidence. The third is the inconsistency between overestimation and overplacement: Those domains that produce the strongest *over*estimation usually produce the greatest *under*placement, and vice versa. We review these three problems below, then we offer a theory that can help resolve them.

Our theory, in brief, is this: People often have imperfect information about their own performances, abilities, or chance of success. But they have even worse information about others. As a result, people's estimates of themselves are regressive, and their estimates of others are even more regressive. Consequently, when performance is high, people will underestimate their own performances, underestimate others even more so, and thus believe that they are better than others. When performance is low, people will overestimate themselves, overestimate others even more so, and thus believe that they are worse than others. We elaborate on and formalize this theory after we first review the empirical inconsistencies the theory is most useful for resolving.

Problem 1: The confounding of overestimation and overprecision

The first problem with overconfidence research is that the most popular research paradigm confounds overestimation and overprecision, making it impossible to distinguish the relative influence of each. This research paradigm measures people's confidence in their correctness at the item level. Of the papers on overestimation, roughly 74% of them use this item-confidence paradigm.

Participants in these studies are asked to report their confidence (usually the probability) that they got a specific problem right. For example, Fischhoff, Slovic, and Lichtenstein (1977) asked their participants general-knowledge questions such as: "absinthe is (a) a liqueur or (b) a precious stone." Participants were then asked to estimate the probability (from 50 to 100%) that they had answered the question correctly. Their confidence systematically exceeded their accuracy. In this paradigm, overestimation and overprecision are one and the same, since being excessively sure you got the item right reflects both overestimation of your performance and excessive confidence in the precision of your knowledge. In order for researchers to distinguish overprecision from overconfidence, it must be possible for respondents to display the two independently. For instance, it must be possible for respondents to be too sure they did worse than they did, combining overprecision with underestimation. In the two-option probabilistic paradigm, however, this wouldn't make any sense. To be excessively sure you incorrectly guessed that absinthe is a precious stone implies you are excessively sure that you know that absinthe is, in fact, a liqueur (Liberman, 2004). One alternative to this item-confidence paradigm is measuring perceptions of performance over a set of items, as we do in the study we present, because it can reduce the confound between overprecision and overestimation.

Problem 2: Underconfidence

The second problem with overconfidence is that the literature documents notable instances of underestimation and underplacement. Underprecision is less common. We consider the evidence on each type of underconfidence in turn.

Underestimation

Underestimation of performance is most likely to occur on easy tasks, on easy items, when success is likely, or when the individual making the estimate is especially skilled (Burson, Larrick, & Klayman, 2006; Fu, Koutstaal, Fu, Poon, & Cleare, 2005; Griffin & Tversky, 1992; Kirchler & Maciejovsky, 2002; Klayman et al., 1999; Koriat, Sheffer, & Ma'ayan, 2002; Krueger & Mueller, 2002; Kruger & Dunning, 1999; Lichtenstein & Fischhoff, 1977; Lichtenstein, Fischhoff, & Phillips, 1982; May, 1986; Sniezek, 1990; Stankov & Crawford, 1997). Erev, Wallsten, and Budescu (1994) pointed out that when people have imperfect knowledge of their own performances, the error in their estimates will make those estimates regressive. As a result, people underestimate their performance when it is high (Burson et al., 2006; Krueger & Mueller, 2002). Underestimation has also been documented in domains other than beliefs about one's own performance, including the illusion of control, the planning fallacy, and optimism about future events.

- *The illusion of control*. As with performance, overestimation of control is most likely to occur when control is low (for reviews see Presson & Benassi, 1996; Thompson, Armstrong, & Thomas, 1998). Studies find that people actually tend to underestimate their control when control is high (Alloy & Abramson, 1979; Jenkins & Ward, 1965).

- *The planning fallacy*. When a task doesn't take long to complete (i.e., the task is easy), people are more likely to overestimate completion times—thereby underestimating their performances (Boltz, Kupperman, & Dunne, 1998; Burt & Kemp, 1994).
- *Pessimism about the future.* People overestimate low-probability risks, including the risk of being injured in a terrorist attack, the risk of getting cancer, or the risk of dying in the coming year (Fischhoff, Bruine de Bruin, Parker, Milstein, & Halpern-Felsher, 2006; Lerner, Gonzalez, Small, & Fischhoff, 2003; Woloshin, Schwartz, Black, & Welch, 1999). Note that these are easy tasks, in the sense that low risk means that the positive outcome is the one most likely to occur.

Underplacement

It is also common to find situations in which people exhibit apparent underconfidence by underplacing their performances relative to others. Underplacement occurs on difficult tasks, in the face of shared burdens, and for rare positive events.

- *Difficult tasks*. On difficult tasks, such as juggling or unicycle riding, people routinely rate themselves below the median (Kruger, 1999; Moore & Small, 2007). In difficult competitions, the contestants become generally pessimistic about winning. In one study, the average participant reported just a 6% chance that they would beat a fellow college student in a quiz on the topic of indigenous vegetation of the Amazon (Windschitl, Kruger, & Simms, 2003).
- *Shared burdens*. When a task becomes more difficult—even if everyone shares the same burden—contestants become more pessimistic about beating the other side (Krizan &

Windschitl, 2007). When negotiators are given a deadline, both the buyer and the seller believe that the time pressure will hurt them and help the other side in the negotiation (Moore, 2004a, 2004b, 2005).

- *Comparative pessimism*. While people believe that common events (even undesirable ones) are more likely to happen to them than to others, they believe that rare events (even desirable ones) are less likely to happen to them than to others (Chambers, Windschitl, & Suls, 2003; Kruger & Burrus, 2004).

Underprecision

Some research has found that the degree of overprecision is sensitive to exactly how the question is asked of participants, with some methods producing less overprecision than others (Budescu & Du, in press; Juslin, Wennerholm, & Olsson, 1999; Juslin, Winman, & Hansson, in press; Winman, Hansson, & Juslin, 2004). These variations, however, do not produce underprecision—only variability in degree of overprecision.

Problem 3: Inconsistency between overestimation and overplacement

The third problem with overconfidence is one that should be apparent after the preceding discussion—its inconsistency. Easy tasks, which produce the most underestimation, also produce the most overplacement. Hard tasks, which produce the most overestimation, also produce the most underplacement (Larrick et al., 2007; Moore & Small, 2007). Research that finds overestimation has tended to focus on difficult domains, such as challenging trivia questions (Campbell, Goodie, & Foster, 2004; Fischhoff et al., 1977; Hoffrage, 2004). Research presenting the most impressive findings of overplacement has tended to focus on easier domains, such as driving a car or getting along with others (College Board, 1976-1977; Svenson, 1981).

In this paper, we attempt to address the three problems identified above, both theoretically and empirically. The theory we present describes the process by which people arrive at beliefs about their own and others' performances, and it can account for the inconsistencies between overestimation and overplacement. We also relate the two of them to overprecision. Finally, we test key aspects of the theory using a novel experiment.

A Theory of Confidence

Stated simply, our theory is this: After experiencing a task, people often have imperfect information about their own performances, but even worse information about the performances of others. As a result, people's post-task estimates of themselves are regressive, and their estimates of others are even more regressive. Imperfect estimates of self and others are regressive for many of the same reasons that imperfect memories produce judgments of objects that revert to the prototype of the group or category from which the object comes (Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000). Consequently, when performance is exceptionally high, people will underestimate their own performances, underestimate others even more so, and thus believe that they are better than others. When performance is low, people will overestimate themselves, overestimate others even more so, and thus believe that they are worse than others. The theory's predictions are illustrated in Figure 1.

In what follows, we elaborate and formalize this theory. Our theory assumes that people hold a subjective probability distribution (SPD) of their own score and an SPD of others' scores over the range of possible outcomes on any task. Overestimation occurs when a person's SPD of his or her own score has a mean that is greater than the person's actual score on the task. Overplacement occurs when a person's SPD of his or her own score has a mean that is greater than the mean of the SPD of others' scores. Overprecision describes an SPD that is narrower (has lower variance) than reality suggests it ought to be.

The Basic Model

For simplicity, we describe our model as it applies to beliefs about scores on a quiz, though the application of the model to additional domains is straightforward. We suppose that a quiz taker, denoted by the index i, has some belief about her likely score. This belief is denoted by the random variable X_i which has some probability distribution over the range of possible scores. We assume that X_i can be decomposed into two components: the global average score on the quiz, denoted by S, and her own, idiosyncratic performance on the quiz, denoted by L_i. We think of S as a proxy for the overall 'simplicity' of the quiz, and L_i as the individual quiz taker's skill relative to others or luck. Both S and L_i are random variables and therefore have their own probability distributions. Thus, $X_i = S + L_i$.

Before seeing the contents of the quiz, the individual has some prior beliefs over the distribution of values that S and L_i can take. As an illustrative example, we might assume that S is normally distributed with mean m_S and variance v_S and that L_i is normally distributed with mean zero and variance v_L. If L_i has a mean of zero, people do not, prior to taking the quiz, expect to be better or worse than others. However, the assumption of no baseline over- or underplacement is not strictly necessary.[2] Given these assumptions, person i believes, *a priori*, that her score X_i will be drawn from a normal distribution with mean m_S and variance v_S+v_L.

When asked about others' scores, we assume person i believes that person j will earn a score $X_j = S + L_j$, where L_j is also a mean-zero normally-distributed random variable with variance v_L. Thus, she believes that her score (X_i) and others' scores (X_j) are random variables with identical distributions.[3]

We consider four points in time, or states of knowledge: Before the quiz, when the individual has no useful information about anyone's performance—either her own or others' (the prior stage), after taking the quiz but before learning her own score or the score of anyone (the interim stage), after learning her own score, but not the scores of others (the posterior stage), and finally, after learning the scores of others (resolution).

At the interim stage, after taking the quiz but before learning her true score, we think of person *i* as having observed a 'signal' of her true score, in the same way that a student taking the SATs has a 'gut feeling' about her performance but won't know her true score for several days or weeks. In our model, *i*'s signal is a realization of the random variable $Y_i = X_i + E_i$, where E_i is a zero-mean random 'error term' that represents the imperfections in the individual's gut feeling. We assume that *i* believes E_i to have a normal distribution with mean zero (her signal is, on average, accurate) and variance v_E.

Overestimation

Suppose person *i* has just completed the quiz but does not yet know her score. She has, however, received a signal y_i, which she believes to be a realization from the random variable Y_i. Knowing y_i allows her to update her belief distribution about her score to $X_i|y_i$, which is her belief about X_i conditional on knowing that Y_i equals y_i. Given the assumptions above, her expected score, $E[X_i|y_i]$, is given by the formula

$$E[X_i|y_i] = \alpha\, m_S + (1- \alpha)\, y_i,$$

where

$$\alpha = (v_L + v_E) / (v_S + v_L + v_E)$$

is between zero and one.[4] In other words, *i*'s expectation after observing her signal (y_i) is that her score lies somewhere between the prior mean (m_S) and the observed signal (y_i).

Assuming *i*'s signal is unbiased, we know that, on average, the signal y_i is equal to the true score x_i. Thus, *i*'s expectation of her own score lies between the prior mean (m_S) and her true score (x_i). If the quiz is easier than expected, then on average, the true score is above the prior mean and therefore *i*'s expectation of her own score is below her true score ($m_S < E[X_i|y_i] < x_i$). Thus, we observe underestimation after easy quizzes. If the quiz is hard then, on average, the true score is below the prior mean and therefore her expectation of her own score is above her true score ($x_i < E[X_i|y_i] < m_S$), implying overestimation after hard quizzes.

While our theory predicts that these results will hold on average, there will of course be exceptions. It is possible, for example, that a person with a high score on an easy quiz could receive an especially high signal y_i, leading to a situation where $m_S < x_i < E[X_i|y_i] < y_i$, which is consistent with overestimation, not underestimation. Because we do not observe the signal y_i in practice, we cannot say much about whether individuals are updating their beliefs optimally, although we can make stronger statements about average beliefs.

Note that our theory's predictions follow directly from Bayes's Law. The intuition behind them is fairly simple: If a person receives a surprisingly high signal of her performance ("Did I just get everything right on that test?"), she discounts that signal because it is likely that at least part of that high signal was due to a positive error in the signal itself ("There must have been some trick questions in there"). From the individual's point of view, this is the best unbiased estimate of her own score. From the researcher's point of view, however, the true score is known, and the individual's estimate appears regressive (on average) because of the difference in information between the two perspectives.

Overplacement

Mathematically, the argument for overplacement is similar to that of overestimation, though the effect runs in the opposite direction. Specifically, suppose now that individual *i* is told her true score x_i and is asked to report her expectation about individual *j*'s score, which she does not yet know. To make such an inference, she must first derive updated beliefs regarding the overall average score on the quiz (the variable *S*), using what she knows about her own performance. We rely on abundant evidence showing that people use themselves as a useful, albeit imperfect basis for predicting others (Krueger, 2000; Krueger, Acevedo, & Robbins, 2005).

Using Bayes's Law, *i*'s expectation of *S* given x_i is

$$E[S|x_i] = \beta\, m_S + (1- \beta)\, x_i,$$

where

$$\beta = v_L / (v_S + v_L)$$

is between zero and one. In other words, *i* believes that the overall average score will lie between her prior expectation for the overall average (m_S) and her own score (x_i). Recall that *i* believes that $X_j = S + L_j$, but as long as *i*'s score is not informative about *j*'s idiosyncratic component to his score (L_j), *i*'s belief is that $E[X_j] = E[S]$. Thus, *i*'s expectation is that *j*'s score lies between her prior expected average score (m_S) and her own score (x_i). If her own score is above the prior expected average (because the quiz was easier than expected), then she exhibits overplacement ($m_S < E[X_j|x_i] < x_i$), and if her own score is below the prior expected average (because the quiz was difficult), then she exhibits underplacement ($x_i < E[X_j|x_i] < m_S$).

The intuition for this result is similar to that for overestimation: If *i* earns a high score, she believes this is partly because the quiz was easier than expected and partly because the idiosyncratic component of her score was positive ("I really kicked that test's butt"). Since the idiosyncratic component of her score is the only thing that differentiates her score from others, she concludes that she outperformed others, on average ("Others probably didn't kick quite as much butt as I did"). Note that this argument holds regardless of whether *i* observes her score (x_i) or only an unbiased signal of her score (y_i). The theory's predictions are illustrated in Figure 1, which shows how expected score by self and others should react to quizzes of varying difficulty (along the x-axis), given a prior (m_S) of five out of ten on the ten-item quiz.

Overprecision

Our theory only predicts errors in estimation and placement when these things are not known precisely. Precision in estimates therefore ought to affect the degree to which people make errors in estimation and placement. Specifically, more precise estimates of self ought to be associated with less over- and underestimation.[5] This is because accuracy ought to affect each, producing both more precise estimates and less error. Likewise, more precise estimates of others ought to be associated with less over- and underplacement.

What is the logic underlying these predictions? If a quiz-taker observed her own score with zero error ($E_i = 0$) and she knew it, then her belief should update to the signal of her score ($E[X_i|y_i] = y_i$), which would be the same as her actual score ($y_i = x_i$). Her beliefs would not, on average, show any over- or underestimation. However, when people receive information that is imperfect and they know it, they should make estimations with wider confidence intervals and thus tend to regress toward their priors. Estimations made by people whose signals are less informative about their true scores ought to be more regressive and ought therefore to display greater underestimation on easy tasks, as well as more overestimation on difficult tasks.

The effect of precision of beliefs about others on overplacement follows similarly. Those who are most unsure about others' performances ought to make the least precise estimates of others' scores and make estimates of others that are most regressive toward their prior expectations. These are the people who are most likely to underestimate others' performances on easy tasks, resulting in overplacement; and most likely to overestimate others on hard tasks, resulting in underplacement.

As to the question of when we should expect overprecision, our theory has little to say.

A Disclaimer

Our theory is based on a model of Bayesian belief-updating processes. However, we should hasten to add that we do not believe that people are perfectly rational Bayesians. Although judgments do sometimes adhere to Bayesian principles remarkably well (Ajzen, 1977; Griffiths & Tenenbaum, 2006; Kersten, Mamassian, & Yuille, 2004), human judgment certainly does not obey Bayes's Law perfectly (Edwards, 1968; Grether, 1990; McKelvey & Page, 1990). The important thing, however, is that while people are not perfect Bayesians, their judgments adhere to the basic logic of Bayes's Law: Their posterior judgments generally fall between their priors and the new evidence. Since our theory's predictions rely only on the direction in which individuals update their beliefs and not on the exact magnitude, the theory's predictions are robust to non-Bayesian beliefs as long as the actual updating process respects the crudest approximation of Bayes's Law.

We have chosen to base the formalization of our theory on Bayesian reasoning not because we believe people always reason in perfect Bayesian fashion, but because it provides a clearer, less ambiguous, and more parsimonious articulation of our theory than alternative approaches could have provided, while providing a reasonable approximation of the psychological processes involved. It is fairly common for researchers to model judgment as Bayesian in this way without necessarily assuming that people comply faithfully with the letter of Bayes's Law (e.g., Windschitl & Wells, 1997).

Our Theory's Relation to Prior Work

Our theory is built on the foundations laid by work that has preceded ours, but makes two key contributions. First, we distinguish between the three different types of overconfidence and specify their relationships to one another. Our second contribution is the application of the principle of updating from prior beliefs based on data, inspired by Bayesian logic. Below we enumerate how our theory is related to, and also distinct from, the most important theories that have offered to explain overconfidence.

Thurstonian Theories of Overconfidence

Our theory is quite consistent with Thurstonian explanations for the hard/easy effect in overestimation, (Erev et al., 1994; Pfeifer, 1994), which rest on the assumption that people make imperfect estimates of their own performances. However, ours is not a straight Thurstonian account. The central role our theory gives to prior expectations and their updating makes it more of a Bayesian than a Thurstonian theory. This is useful for explaining regressiveness in performance estimates that are not constrained by floor or ceiling effects (Moore & Small, in press). Perhaps more importantly, Thurstonian theories have not yet been applied to explaining overplacement, whereas our theory offers such an explanation.

Brunswikian Theories of Overconfidence

Ecological validity

Some important explanations for overconfidence have been inspired by Brunswik's (1952) concern for ecological validity in judgmental tasks. Brunswikian explanations for overconfidence hold that overestimation of task performance is an artifact of the way items were selected. The solution to such selection is to establish a reference class with which people are likely to be familiar, and randomly select judgments from that set (Gigerenzer, Hoffrage, & Kleinbölting, 1991). Our theory would offer two explanations for the tendency for overconfidence to go down when problems are randomly drawn from a familiar reference class. First, such random selections will necessarily produce some very simple items. These simple items are more likely to produce underestimation than overestimation, consistent with our theory (and with the work of Soll, 1996). Our theory is clearly consistent with the ecological explanation in suggesting that prior research finding overestimation of task performance may have used tasks that were more difficult than participants expected them to be. Second, the more familiar the class of problems is, the better information people have about their performances. If, in familiar domains, people are better at determining whether they are right or wrong, then we should expect to see less over- or underestimation of performance.

The BIAS model

Fiedler’s (1996) Brunswikian induction algorithm for social cognition (BIAS) focuses on the superior information people have about themselves and their ingroups rather than other people and outgroups. Our work builds on Fiedler’s work in several ways. First, we show how this same basic theoretical model can account for findings of underplacement on difficult tasks. Second, our theory elaborates on Fiedler’s model by explicitly allowing for the possibility that people use information about themselves to make inferences about others (Moore & Small, 2007). Third and most importantly, while Fiedler’s BIAS model is not inconsistent with the possibility that agents use the information at their disposal to update their prior beliefs, Fiedler does not develop this Bayesian logic. This omission has important repercussions because the predictions of Fiedler's theory will not hold under some prior beliefs. The development of the Bayesian logic is essential for explaining the overconfidence effects our theory seeks to account for. It is also crucial for generalizing the explanation to circumstances in which signals (and beliefs) are not constrained by ceilings or floors, such as judgments of profitability, size, or time to completion.

Hypotheses

Our theory makes four basic predictions regarding overestimation and overplacement:

Overestimation: On difficult tasks, people will overestimate their performances on average.

Underestimation: On easy tasks, people will underestimate their performances on average.

Overplacment: On easy tasks, people will, on average, overplace their performances relative to the performance of others.

Underplacement: On difficult tasks, people will, on average, underplace their performances relative to the performance of others.

Returning to our example of students taking a quiz, when the quiz is more difficult than expected, our theory predicts that people will overestimate but underplace their scores. When the quiz is easier than expected, people will underestimate but overplace their scores. We expect all four of these effects to follow straightforwardly from updating prior beliefs (expectations) using imperfect information about performance. In our experimental paradigm, as in most of life, people have imperfect information about the performances of both self and others, but they have much better information about themselves than they do about others.

Our theory implies that the precision with which people are able to estimate performance should moderate the effects of task difficulty on both estimation and placement. In particular, we expect that those individuals who are least confident in their estimations of their own performances will make estimates that are most regressive toward the prior and will also show the least precision (i.e., the greatest variance). If this effect holds, then the overestimation (on hard tasks) and underestimations (on easy tasks) will be associated with greater precision in estimates of one's own performance. Those who are more accurate and know it will make estimates that are both more precise and less regressive.

Likewise, our theory implies that accuracy in beliefs about others will result in both more precision of estimates and also in less overplacement (on easy tasks) and underplacement (on hard tasks). On the other hand, individuals who are least confident in their estimations of others' performance will make the most regressive estimates of others. If this effect holds, then greater precision in the estimates of others will be associated with less overplacement on easy tasks and underplacement on hard tasks.

Our theory implies that we ought to be pessimistic regarding the possibility that feedback or experience could overcome any of these biases. If they represent the result of Bayesian judgment processes, then there is little room to improve upon them. Our experiment employs a repeated-measures design that allows us to examine the effects of experience.

ILLUSTRATIVE STUDY

Method

Participants

Participants were 82 students from Carnegie Mellon University. Participants were recruited using the following posting on a web site that recruits experimental participants under the name "*Tons o' Trivia*" with the following description: "*Participants in this study will take a number of different trivia quizzes. They will get paid based on their performance and based on their ability to guess how they and others did*." Given that this recruiting method was likely to attract trivia enthusiasts, results obtained by Camerer and Lovallo (1999) suggest it might bias our results in favor of finding overplacement. We do not find evidence of systematic overplacement in our results, and this concern does not offer a compelling alternative explanation for the pattern of over- and underplacment, combined with under-and overestimation we predict and find.

Participants were, on average, 24 years old (*SD* = 6.42, range 18 to 59). Males constituted 51% of participants and females constituted 49%. Thirty-eight percent of participants identified themselves as white, 38% as Asian or Asian-American, 11% as African or African American, 5% as Hispanic, and 8% as another racial category.

Procedure

The entire experiment was run on computers via a web server, but all participants were present in the experimental laboratory.[6] In each of 18 rounds, each participant took a 10-item trivia quiz. The 18 quizzes included three quizzes on each of six topics: science, movies, history, sports, geography, and music.[7] Each topic included one easy, one hard, and one medium difficulty quiz. Each participant encountered a different order of these 18 quizzes. The ordering was random, subject to the constraint that each three-round block include one quiz at each level of difficulty. The result was six three-round blocks.

In each round, participants earned $25*r*, where *r* was his or her percentile rank relative to all other participants who had already taken that quiz. For the sake of computing this percentile rank, participants were counted as having scored better than half and worse than half of those who had obtained the same score.

Before taking the quiz each round, participants were asked to predict the probability (*p*) that they would obtain each of the 11 possible scores, 0 through 10. Predictions were rewarded according the quadratic scoring rule, as follows. Participants earned 1 + r - w dollars, where $r = 2p$ for the score they actually received and w equals the sum total of each p^2 for each of the 11 scores. This quadratic scoring rule maximizes participants' expected payoffs if they accurately report their best estimate of the truth (Selten, 1998). It is equal to $2 when a participant correctly estimates his or her own score with certainty (100% confidence), because $r = 2$ and $w = 1$. It is equal to $0 when a person attaches 100% confidence to a score she did not obtain, because $r = 0$ and $w = 1$. Participants were instructed, "*This formula may appear complicated, but what it means for you is very simple: You get paid the most when you honestly report your best guesses about the likelihood of each of the different possible outcomes. The range of your payoffs is from $0 to $2 for each set of guesses*."

Participants made these predictions on the computer by sliding each of eleven bars to indicate probability. When they slid one bar, the other ten automatically adjusted proportionally so that the total probability across all eleven bars was always equal to 100%. These eleven bars started out with different random settings for each set of predictions a participant made.

In addition to predicting what their own score would be, participants predicted the score of a randomly selected previous participant (the RSPP). Participants made "prior" predictions before they had any specific information about the quiz they were about to take. Then they took the actual quiz. Participants then estimated their own quiz scores[8] and the RSPP's score again—their "interim" estimates. Then they were shown the correct answers to the quiz and were asked to grade themselves. They were then again asked to make a "posterior" estimate the RSPP's score. Finally, they were given full feedback for the round: Their own scores, the RSPP's score, their payoffs for each of the five predictions (two for their own and three for the RSPP's score), and their percentile ranks.

Prior research on overplacement has commonly asked people to compare themselves with a group (such as other students at the same university) or with a typical member of a group (such as an average university student). However, asking people to consider a randomly selected other as we do offers crucial advantages for testing our theory. The alternative, asking individuals to report their beliefs about some summary statistic of the population of other test-takers, leads to fundamental difficulties in applying our theory directly because the variance of the summary statistic (such as the median or mean) will typically be much lower than the variance of an individual's performance. Second, asking people about the RSPP has the nice consequence that it becomes optimal for people to report their subjective probability distribution of the scores of all other participants. Subjective probability distributions of the median, by contrast, ought to have a much smaller variance.

Dependent measures

We derived all our key dependent measures of overestimation, overplacement, and overprecision by comparing participants' score estimations with actual scores. We can compute estimated scores by multiplying each of the eleven possible scores by its reported likelihood, and summing these.

Overestimation. A participant's actual score was subtracted from his or her reported estimated score to measure overestimation.

Overplacement. We computed a measure of overplacement that takes into account whether a participant really is better than others. We used the following formula:

$$Overplacement = (E[X_i] - E[X_j]) - (x_i - x_j)$$

Where $E[X_i]$ is an individual's belief about his or her own expected performance, $E[X_j]$ is that person's beliefs about the expected performance of the RSPP on that quiz, and x_i and x_j refer to the actual scores of the individual and the RSPP. In other words, this measure begins with an individual's belief that he or she is better than others and corrects that for the degree to which he or she actually is better than others.

Overprecision. While prior research has generally measured overprecision by having participants specify confidence interval ranges, we elicit the subjective probability distribution over all possible outcomes. One of the benefits of this approach is that our measures are not subject to the overprecision that arises due to constraints on memory capacity as documented by Hansson, Juslin, and Winman (2006; see also Juslin et al., in press). Our novel approach also allows us to test for overprecision in a variety of ways, as we detail in our presentation of the results.

Payoffs

Five of the 18 quiz rounds were chosen at random as payoff rounds. Participants' earnings in those five rounds were averaged to compute their actual payoffs for the experiment. Participants were paid (an average of $20.23 after about 90 minutes in the laboratory), debriefed, thanked, and dismissed.

Results

Manipulation checks

In order to check that our manipulation of difficulty affected quiz performance as we intended, we submitted participants' quiz scores to a 6 (block) X 3 (difficulty) within-subjects ANOVA. The results show the expected main effect of difficulty, F (2, 162) = 781, $p < .001$, $\eta^2 = .91$. Scores on easy quizzes were higher (M = 8.86) than on medium quizzes (M = 5.92) which were higher than scores on hard quizzes (M = .71).

Participants correctly perceived differences in difficulty between the different quiz types. We analyzed reported scores for the self at the interim phase using a 6 (block) X 3 (difficulty) within-subjects ANOVA. The results show a significant main effect for difficulty, F (2, 162) = 586, $p < .001$, $\eta^2 = .88$. Participants believed they had scored better on quizzes that were easy (M = 8.64) than medium (M = 5.93) or hard (M = 1.50). Clearly, however, the effect of the difficulty manipulation is smaller on beliefs about performance than it is on actual performance. This is consistent with our theory, and gives rise to the well-documented hard/easy effect on estimation.

Note that we do not analyze results from the first round separately because they do not differ from, and are therefore redundant with, the repeated-measures analyses.

Overestimation

Participants' interim overestimation was subject to a 6 (block) X 3 (difficulty) within-subjects ANOVA. As expected, there was a main effect of difficulty, F (2, 162) = 42.53, $p <$ .001, $\eta^2 = .34$. Participants underestimated their performances on easy quizzes (M = -.22), overestimated them on difficult quizzes (M = .79), and were (on average) accurate on medium quizzes (M = .01). There was no main effect of block, F (6, 405) < 1, $p = .89$, but the block X difficulty interaction does attain statistical significance, F (10, 810) = 2.65, $p = .003$, $\eta^2 = .03$. This interaction implies change in the effect of quiz difficulty on overestimation over the six blocks. However, as Table 1 shows, this change is not a linear change in over- or underestimation with time. Tests for linear patterns in the change do not attain statistical significance. In other words, participants are not learning to make fewer errors estimating their scores over the 18 rounds.

Overplacement

At the prior phase, our participants do not, on average, report believing that they (M = 5.36) will perform better than the RSPP (M = 5.36).

Participants' interim overplacement was subject to a 6 (block) X 3 (difficulty) within-subjects ANOVA. There was a main effect of difficulty, F (2, 162) = 24.88, $p < .001$, $\eta^2 = .24$. Participants overplaced their performances on easy quizzes (M = .48), underplaced their performances on difficult quizzes (M = -1.36), and demonstrated no over- or underplacement on medium quizzes (M = .04). There was no main effect of block, F (6, 405) = 1.04, $p = .39$, η^2 = .01, nor a block X difficulty interaction, F (10, 810) < 1, $p = .99$. See Table 2. Remember that these are measures of bias—that is, underplacement is the belief that one is worse than the RSPP, corrected for the degree to which one actually *is* worse than the RSPP.

The relationship between overestimation and overplacement

One of the contributions of this paper is to document the interrelationships between different forms of overconfidence. This is made possible by the fact that we measure all three varieties of overconfidence on the same task.

Within any given quiz, overestimation and overplacement are correlated, with an average correlation coefficient of .29. This is for trivial reasons: both measures are an increasing function of an individual's beliefs about his own score and a decreasing function of his actual score (see Larrick et al., 2007). On the other hand, across the 18 quizzes, overestimation and overplacement are negatively and significantly correlated, r (18) = -.64, $p = .004$. This is a natural consequence of the effects documented above: task ease affects overestimation and overplacment in opposite ways. See Figure 2.

Two features of Figure 2 deserve mention. First, the regressiveness of estimates for both self and the RSPP is obviously smaller for easy than hard quizzes. Second, the linear regression lines for estimates of self and the RSPP do not cross the identity line at 5.36 (participants' prior expected score), but closer to 7. Both of these results may be attributable to the fact that participants had better information about performance on easy quizzes than on difficult ones. Participants are likely to know whether they have correctly answered the easy question, "What is the capital of the United States?" They are less likely to know whether they (or anyone else) has correctly answered the hard question, "What album ranks #1 on the best-selling albums of all time, just ahead of Michael Jackson's 'Thriller'?"[9]

Overprecision

Unlike overestimation and overplacement, participants' judgments *do* appear to show systematic overprecision. We computed, for each individual estimate, the range of scores that came closest to that individual's 90% confidence interval. On average, this interval covered 3.11 of the 11 possible scores, and represented 90.5%[10] confidence. The actual score fell within this range 73.1% of the time. This 90.5% confidence is significantly greater than the 73.1% accuracy; a binomial test confirms that the odds of observing accuracy rates of 73.1% or lower for 90.5% confidence intervals are about 6 per million. Measured this way, confidence exceeded accuracy across all quiz difficulties, all rounds, all phases, and for both self and RSPP, all *p*s < .01. We tested the robustness of these results using several additional tests of overprecision, including: (1) using 50%, 60%, 70%, and 80% confidence intervals; (2) comparing the variance in actual scores with the variance in each subjective probability distribution; and (3) comparing confidence with accuracy for the score at the peak of each subjective probability distribution. These different approaches also produced consistent evidence of overprecision.

Our participants' 73.1% hit rate may appear substantially above prior findings, which documented hit rates of 30% to 50% for 90% confidence intervals (Teigen & Jorgensen, 2005). However, we should note that hit rates are lowest at the prior phase (55.1%) when participants were guessing a quantity about which they were poorly informed, as has been the case in most prior research finding overprecision. Hit rates are substantially higher at the interim (85.6%) and posterior (84.4%) phases, consistent with research showing that knowledge of the judgment domain moderates the strength of overprecision (Block & Harper, 1991). However, even here, accuracy is significantly below confidence.

It is worth noting that confidence and accuracy are nonetheless correlated. The width of a confidence interval at the interim phase is positively correlated with the distance between that estimate and the actual score, both for self (r = .54, p < .001) and the RSPP (r = .23, p < .001). Those who make the least accurate score estimates are also those to provided the broadest confidence intervals.

The relationship between overprecision, overestimation, and overplacement. As our theory predicts, the precision of participants' judgments are related in interesting ways to overestimation and overplacement. These associations are most fruitfully examined at the interim phase, where participants have some information about performance on that quiz and we have measures of beliefs for both self and others.

We predicted that precision in estimates of one's own score would be associated with less underestimation on easy quizzes and also with less overestimation on hard quizzes. Indeed, this was the case. On the easy quizzes, the correlation between confidence interval width (i.e., lack of precision) and *under*estimation is positive (r = .45, p < .001). Participants' tendency to underestimate their scores on easy quizzes was greater among those who reported broad confidence intervals. On the hard quizzes, the correlation between confidence interval width and *over*estimation is also positive ($r = .47$, $p < .001$), because the broader the confidence interval, the more the overestimation. In sum, precise estimates were associated with less over- and underestimation.

We also predicted that greater precision in estimates of the RSPP's score would be associated with less overplacement on easy quizzes and also with less underplacement on hard quizzes. Indeed, this was the case. On easy quizzes, the correlation between confidence interval width (i.e., lack of precision) and *over*placement is positive ($r = .38$, $p < .001$): the broader the confidence interval in estimating others, the more the overplacement. On hard tests, the correlation between confidence interval width and *under*placement is also positive ($r = .27$, $p < .001$) because the broader the confidence interval in estimating the RSPP, the more the underplacement. Precise estimates of the RSPP's score decrease the tendency to underplace one's own performance on hard quizzes. In sum, precise estimates were associated with less over- and underplacement.

Origins of beliefs about self and others

Beyond specifying the relationships between the three different forms of overconfidence, our theory specifies the processes that give rise to post-quiz beliefs about self and others. One of the unique features of our data is that we collected measures of prior beliefs. In order to examine the degree to which participants updated their beliefs about themselves and others, we computed measures of updating by computing the absolute difference from prior to interim in estimated scores. These measures were then subject to a 3 (difficulty) X 6 (block) X 2 (target: self vs. other) within-subjects ANOVA. Participants updated their beliefs from prior to interim more so for self ($M = 3.43$) than for others ($M = 2.63$), resulting in a main within-subjects effect for target, F (1,

81) = 98.23, $p < .001$, $\eta^2 = .55$. This makes sense, since participants at the interim stage had better information about themselves than they did about others. The results also reveal a main effect of difficulty, since participants updated more for the simple ($M = 3.45$) and the difficult ($M = 3.51$) than the medium ($M = 2.13$), $F(2, 162) = 67.86$, $p < .001$, $\eta^2 = .46$. This also is consistent with our theory, since participants' priors were more accurate for the medium than for the easy or difficult quizzes.

Our theory predicts that participants' interim beliefs about their own scores should be a joint function of their prior expectations and the private signals they receive regarding their own performance (i.e., their experience taking the quiz). These predictions are borne out by the regression results presented in Table 3. A few features of these regression results merit comment. First, the prior featured more heavily in interim estimates of the RSPP; estimates of others are more regressive. Second, participants' own actual scores featured more heavily in estimations of their own scores; they had better information about themselves than others, and knew it. Third, quiz ease was a useful predictor of beliefs about others, but not for self (after controlling for the individual's own actual score). People tried to infer how difficult the test would be for others when making predictions about the RSPP, but for themselves, they could rely directly on how well they had done on the quiz. These process results are all consistent with our theory.

DISCUSSION

Our theory posits a parsimonious explanation, based on basic Bayesian principles of belief-updating, for the negative relationship between overestimation and overplacement across tasks. Our results are consistent with the predictions of this theory, which begins with the assumption that people have imperfect knowledge of their own performances and even more imperfect knowledge of others' performances. As a result of this imperfect knowledge, beliefs tend to regress toward prior expectations (more so for others than for self), producing the pattern of results we observe.

Our experiment allowed us to measure all three varieties of overconfidence so that we could relate them to each other and to the predictions of our theory. We find that overestimation *increases* with task difficulty but that overplacement *decreases* with task difficulty. As a result, estimation and placement are negatively correlated across tasks. While our theory cannot explain the persistent presence of overprecision in our data, our theory can help explain the effect of precision on estimation and placement. Consistent with our theory, greater precision in self-estimates is associated with both reduced over- and underestimation. In addition, greater precision in other-estimates is associated with both reduced over- and underplacement.

Implications

It should be clear that our theory applies not just to performance (as in our experiment), but would generalize to risk perception and to inferences about abilities, behaviors, and traits, both retrospectively and prospectively.

Extensions and applications of our theory

Our theory can help explain a number of otherwise incongruous research findings. For instance, evidence shows that the risks to which people believe that they are less vulnerable than others are often those where they most overestimate their own actual risk. For instance, women dramatically overestimate their chances of contracting breast cancer yet believe they are at less risk than other women (Woloshin et al., 1999). Americans overestimate the risk of being injured in a terrorist attack but believe they are at less risk than other Americans (Lerner et al., 2003). Chinese respondents reported that they were at less risk than their peers for contracting SARS, but they nevertheless overestimated their actual chances of catching the disease (Ji, Zhang, Usborne, & Guan, 2004). People overestimate their risk of dying in the coming year (Fischhoff et al., 2006) yet believe they are at less risk of death than their peers (Weinstein, 1980). And most notably, smokers overestimate their chances of getting lung cancer (Viscusi, 1990) but believe they are at less risk than other smokers (Slovic, 2001).

Our theory may be able to reconcile these findings. Take, for example, the the Slovic-Viscusi debate about whether smokers over- or underestimate their risk. Smokers do indeed overestimate the probability of contracting lung cancer, as Viscusi (1990) shows. But they also believe that others are more likely than they are to experience this rare event, as Weinstein (1998) shows. Our theory can accommodate these findings if people have imperfect information about their own risky behavior and its consequences, but have better information about their own behavior than that of others. Likewise, when it comes to predicting whether they are going to die in the coming year, people tend to overestimate this small probability. Nevertheless, because people have better information for themselves than for others, they overestimate others more than self and consequently believe they are at less risk than others.

Can Our Theory Accommodate Other Moderators of Overconfidence?

Prior research has documented a number of important moderators of overconfidence. Below we discuss ways in which our theory is (and is not) compatible with these prior findings.

Controllability. A number of studies have found that the tendency to believe that one's performance will be better than that of others is stronger for outcomes under personal control than for chance outcomes (for reviews see Harris, 1996; Klein & Helweg-Larsen, 2002). Our theory would explain such effects by pointing out that when performance is entirely determined by chance, it is not possible for people to have better information about their own future performances than about others', because both will be determined for them by Lady Luck. Our theory would predict, however, that controllability will lead people to believe that they are *worse* than others on hard tasks. This prediction has been confirmed by Moore and Cain (2007). Their results showed that people believe they are better than others at easy tasks, worse than others on hard tasks, and equal to others when performance is determined by chance.

Observability. People believe that they are more likely than others to exhibit common unobservable behaviors (such as honesty); however, this belief is reduced for common observable behaviors (such as friendliness) (Alicke, 1985; Allison, Messick, & Goethals, 1989; Paunonen, 1989). As our theory would predict, this effect reverses itself for rare behaviors (Kruger & Savitsky, 2006). Our theory's explanation for these results is simply that people have worse information about others on unobservable than observable traits and behaviors. Differential regressiveness to the prior occurs when people have better information about themselves than others. The lack of information about others for unobservable behaviors makes people's estimates of others more regressive, exaggerating the tendency to believe that one is more likely than others to engage in common behaviors and less likely than others to engage in rare behaviors.

Frequentistic vs. probabilistic judgment. Prior research has found more overconfidence on probabilistic than frequentistic estimations of performance (Gigerenzer et al., 1991). However, these probabilistic judgments have usually come from the problematic item-confidence probabilistic paradigm, and so it is possible that the difference is accounted for by the role of overprecision in that paradigm. It is possible to measure perceived performance at the item level without reliance on probabilities. For instance, runners can estimate their race times, or test-takers can estimate the distance of their own answers from the correct answers. We would expect to find little or no systematic overestimation of performance for these sorts of judgments, after controlling for task difficulty.

Item selection. Test items randomly selected from within a reference class are likely to include a mixture of easy and hard items. This is one of the reasons why such representative sets produce less overconfidence (Soll, 1996). We should note that it is for this reason that we could not use randomly selected items to constitute the trivia questions in our study. Random selection from a given class will necessarily mix the difficulty levels of the questions, whereas we wanted to manipulate difficulty systematically.

Personal experience. Some research has found that personal familiarity with some event or outcome seems to increase comparative optimism (Weinstein, 1980). The findings on familiarity contrast with those of observability: Observability increases information about performance or outcomes, particularly for others. Personal familiarity with one's own chances of experiencing an event, on the other hand, is likely to increase the disparity between the quality of information about self vs. others. The fact that personal experience should exacerbate overplacement for common events, then, is consistent with our theory.

Training and improvement. Driving an automobile is a complex task that is certainly difficult when we start doing it. But we get better at it and it gets easier. Our theory is consistent with the evidence which says that beginning drivers (for whom the task is difficult) think they're worse than average (Rutter, Quine, & Albery, 1998). Experienced drivers think they're better than average (Svenson, 1981). When they're learning a task, people overestimate their performance. As they gain experience, they are more likely to underestimate their performance (Koriat et al., 2002) but come to believe that they are better than others.

Cognitive vs. perceptual judgments. Some prior research has suggested that tendencies toward overconfidence are stronger in cognitive than in perceptual judgments (Keren, 1988). Björkman, Juslin, and Winman (1993) asked their participants to determine which of two weights was heavier and which of two rectangles, viewed from a distance, was longer. Participants then made confidence judgments regarding their answers. The results suggested a general tendency toward underconfidence in these judgments. However, other researchers replicate the hard/easy effect, wherein hard tasks generate overconfidence and easy tasks generate underconfidence, for both cognitive and perceptual judgments (Baranski & Petrusic, 1999). Unfortunately, all of these studies use the item-confidence paradigm that makes it impossible to distinguish the roles of overestimation and overprecision. More importantly, none of these studies measure participants' expectations for performance. It is possible that there are systematic differences between cognitive and perceptual judgments with respect to prior expected difficulty, and that this can account for some of the observed differences.

Direct vs. indirect measures. The leading prior explanation for above- and below-average effects (over- and underplacement) is differential weighting. This explanation holds that when making comparative judgments, people overweight self-knowledge and underweight knowledge about others (Chambers et al., 2003; Giladi & Klar, 2002; Kruger, 1999). In our study we only measured comparative judgments indirectly. We did not directly ask people to compare themselves with others. Research has generally found that both overplacement on easy tasks and underplacement on difficult tasks is stronger when comparisons are elicited directly than indirectly (Chambers & Windschitl, 2004). Our theory cannot account for this difference in elicitation formats.

However, some research suggests that discrepancies between direct and indirect elicitation methods are an artifact of question format, and that parallel question formats can eliminate the discrepancy (Burson & Klayman, 2005; Moore, 2007). And while the differential weighting explanation is useful for accounting for some important results in direct comparative judgments (Kruger, Windschitl, Burrus, Fessel, & Chambers, in press), it cannot account for the presence of overplacement or underplacement effects in indirect comparative judgments, such as those we present. (Indirect comparative judgments are computed using the implicit comparison between individual judgments of self and other.)

Extensions to Other Phenomena

Fiedler's work (1991; 1996; 2000) has persuasively demonstrated the broad applicability of explanations such as the one we present in this paper. The Bayesian logic presented by our theory is useful for specifying the conditions under which the effects hypothesized by Fiedler's theory will occur and when they will not. In particular, Fiedler's BIAS model hypothesizes that when we receive predominantly positive information about two groups, we will wind up with a more positive impression of the group about whom we get more information. This is, for instance, why people have more favorable attitudes toward ingroups and why outgroups are perceived as more homogeneous (Sanbonmatsu, Sherman, & Hamilton, 1987). Our theory helps clarify the question of "Positive relative to what?" The answer is that information must be, on balance, better than prior expectations for it to influence judgments. When people have positive impressions of an ingroup and outgroup to begin with (before they gain more information about the ingroup), additional positive information will be uninformative, and the predictions of the BIAS model will not hold.

Limitations

The new data we present are illustrative, not definitive. We cannot claim that the pattern we observe would hold regardless of context, task domain, or subject population, although our theory does not suggest that these factors should matter. Our theory is, however, not meant to account for all overconfidence effects. For instance, our theory cannot explain motivational effects on overconfidence, nor can it account for the role of elicitation formats.

Motivational effects. Our theory cannot account for evidence showing that estimations of future performance is reduced when the time for the task draws near (Carroll, Sweeny, & Shepperd, 2006; Gilovich, Kerr, & Medvec, 1993; Nussbaum, Liberman, & Trope, 2006). Nor can we account for evidence showing that estimations of future performance go down when the task is real vs. hypothetical (Armor & Sackett, 2006). However, we ought to point out that there are situations in which people may be motivated to report that they are worse than others. Such is the case with social status, where overestimating one's status can lead to ostracism and intra-group conflict (Anderson, Srivastava, Beer, Spataro, & Chatman, 2006). And while there are some domains in which overconfidence can be adaptive, it can also undermine effort and performance (Stone, 1994; Vancouver, Thompson, Tischner, & Putka, 2002) and it can lead to greater disappointment when performance falls short of inflated expectations (McGraw, Mellers, & Ritov, 2004).

Motivational effects on confidence judgments are real and robust. They are just outside the purview of our theory. And we would hasten to note that none of the theories that can account for these motivational effects can account for the empirical results we present: underestimation on easy tasks, underplacement on difficult tasks, and overprecision's reduction of both overerstimation and overplacement.

Individual differences. While not every study that looks has found a relationship (Jonsson & Allwood, 2003), some prior research has found overconfidence to be related to individual differences, including personality (Schaefer, Williams, Goodie, & Campbell, 2004), gender (Niederle & Vesterlund, 2007; Pulford & Colman, 1997; Soll & Klayman, 2004), narcissism (Ames & Kammrath, 2004; Campbell et al., 2004), and cognitive abilities (Kleitman & Stankov, in press; Stankov & Crawford, 1996). While our theory cannot explain these prior results, we nevertheless wanted to examine the effects of these individual differences on all three varieties of overconfidence, so our study included measures of all these individual differences. We do not present those results in this paper because we did not find any significant correlations between any of the individual differences we measured and any of the three varieties of overconfidence, so those results are not particularly interesting.

Conclusions

We began this paper by discussing findings on overconfidence that seem inconsistent. Part of the reason for the apparent inconsistency was that overconfidence has been studied in three distinct ways, and prior research has not always been good about distinguishing them. The most common paradigm used to study overconfidence actually confounds two varieties—overestimation and overprecision—with one another.

In this paper, we have attempted to present an explanation for overconfidence that can help relate the three different varieties to each other and explain their inconsistencies in empirical evidence. Most notably, our theory explains the negative relationship between overestimation and overplacement across tasks. It also leads to the prediction, confirmed in the data we present, that more precise beliefs tend to be associated with both less overestimation and less overplacement. The evidence we have reviewed and the new evidence we present both suggest that overestimation, overplacement, and overprecision are *not* different manifestations of the same underlying construct. The three different types of overconfidence are conceptually and empirically distinct.

References

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology, 35*(5), 303-314.

Alba, J. W., & Hutchinson, J. W. (2000). Knowledge calibration: What consumers know and what they think they know. *Journal of Consumer Research, 27*, 123-156.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49*(6), 1621-1630.

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

Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter than others: The Muhammad Ali effect. *Social Cognition, 7*(3), 275-295.

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General, 108*, 441-485.

Alpert, M., & Raiffa, H. (1969/1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Ames, D. R., & Kammrath, L. (2004). Mind-reading and metacognition: Narcissism, not actual competence, predicts self-estimated ability. *Journal of Nonverbal Behavior, 28*, 187-209.

Anderson, C., Srivastava, S., Beer, J. S., Spataro, S. E., & Chatman, J. A. (2006). Knowing your place: Self-perceptions of status in face-to-face groups. *Journal of Personality and Social Psychology, 91*(6), 1094-1110.

Armor, D. A., & Sackett, A. M. (2006). Accuracy, Error, and Bias in Predictions for Real Versus Hypothetical Events. *Journal of Personality and Social Psychology, 91*(4), 583-600.

Baranski, J. V., & Petrusic, W. M. (1999). Realism of confidence in sensory discrimination. *Perception & Psychophysics, 61*(7), 1369-1383.

Barber, B., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics, 116*(1), 261-293.

Belsky, G., & Gilovich, T. (2000). *Why smart people make big money mistakes--and how to correct them: Lessons from the new science of behavioral economics* (1st Fireside ed.). New York, NY: Simon & Schuster.

Berger, J. O. (1980). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York: Springer-Verlag.

Björkman, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception and Psychophysics, 54*(1), 75-81.

Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. *Organizational Behavior and Human Decision Processes, 49*(2), 188-207.

Boltz, M. G., Kupperman, C., & Dunne, J. (1998). The role of learning in remembered duration. *Memory & Cognition, 26*(5), 903-921.

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: Chicago University Press.

Budescu, D. V., & Du, N. (in press). The coherence and consistency of investors' probability judgments. *Management Science*.

Burson, K. A., & Klayman, J. (2005). *Judgments of performance: The relative, the absolute, and the in-between*. Ann Arbor: Unpublished manuscript. Available at SSRN: http://ssrn.com/abstract=894129

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90*(1), 60-77.

Burt, C. D. B., & Kemp, S. (1994). Construction of activity duration and time management potential. *Applied Cognitive Psychology, 8*(2), 155-168.

Camerer, C. F., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review, 89*(1), 306-318.

Campbell, W. K., Goodie, A. S., & Foster, J. D. (2004). Narcissism, confidence, and risk attitude. *Journal of Behavioral Decision Making, 17*(4), 481-502.

Carroll, P., Sweeny, K., & Shepperd, J. A. (2006). Forsaking optimism. *Review of General Psychology, 10*(1), 56-73.

Chambers, C., & Healy, P. J. (2007). *Updating towards a signal*. Pasadena, CA: Unpublished manuscript.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivational factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130*(5).

Chambers, J. R., Windschitl, P. D., & Suls, J. (2003). Egocentrism, event frequency, and comparative optimism: When what happens frequently is "more likely to happen to me". *Personality and Social Psychology Bulletin, 29*(11), 1343-1356.

College Board. (1976-1977). *Student descriptive questionnaire*. Princeton, NJ: Educational Testing Service.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance, 53*(6), 1839-1885.

Diaconis, P., & Ylvisaker, D. (1979). Conjugate priors for exponential families. *Annals of Statistics, 7*(2), 269-281.

Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17-52). New York: Wiley.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101*(3), 519-527.

Fiedler, K. (1991). The tricky nature of skewed frequency tables: An information loss account of distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology, 60*(1), 24-36.

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review, 103*(1), 193-214.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review, 107*(4), 659-676.

Fischhoff, B., Bruine de Bruin, W., Parker, A. M., Milstein, S. G., & Halpern-Felsher, B. L. (2006). *Adolescents' perceived risk of dying.* Paper presented at the Paper presented at the Center for Risk Perception and Communication Seminar Series, Carnegie Mellon University.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*(4), 552-564.

Fu, T., Koutstaal, W., Fu, C. H. Y., Poon, L., & Cleare, A. J. (2005). Depression, Confidence, and Decision: Evidence Against Depressive Realism. *Journal of Psychopathology and Behavioral Assessment, 27*(4), 243-252.

García, D., Sangiorgi, F., & Urosevic, B. (2007). Overconfidence and market efficiency with heterogeneous agents. *Economic Theory, 30*(2), 313-339.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*(4), 506-528.

Giladi, E. E., & Klar, Y. (2002). When standards are wide of the mark: Nonselective superiority and inferiority biases in comparative judgments of objects and concepts. *Journal of Experimental Psychology: General, 131*(4), 538-551.

Gilovich, T., Kerr, M., & Medvec, V. H. (1993). Effect of temporal perspective on subjective confidence. *Journal of Personality and Social Psychology, 64*(4), 552-560.

Glaser, M., & Weber, M. (2007). Overconfidence and trading volume. *Geneva Risk and Insurance Review, 32*, 1-36.

Grether, D. M. (1990). Testing Bayes rule and the represetativeness heuristic: Some experimental evidence. *Journal of Economic Behavior and Organization, 17*, 31-57.

Griffin, D. W., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*(3), 411-435.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17*(9), 767-773.

Hansson, P., Juslin, P., & Winman, A. (2006). *The role of short term memory and task experience for overconfidence in judgment under uncertainty*.

Harris, P. (1996). Sufficient grounds for optimism? The relationship between perceived controllability and optimistic bias. *Journal of Social and Clinical Psychology, 15*(1), 9-52.

Healy, P. J., & Moore, D. A. (2007). *Bayesian overconfidence*. Tepper Working Paper 2007-E16.

Hoelzl, E., & Rustichini, A. (2005). Overconfident: Do you put your money on it? *Economic Journal, 115*(503), 305-318.

Hoffrage, U. (2004). Overconfidence. In R. F. Pohl (Ed.), *Cognitive illusions: Fallacies and biases in thinking, judgment, and memory* (pp. 235-254). Hove, UK: Psychology Press.

Howard, M. E. (1983). *The causes of war and other essays*. Cambridge, MA: Harvard University Press.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98*, 352-376.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129*, 220-241.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs, 79*(1), 1-17.

Ji, L.-J., Zhang, Z., Usborne, E., & Guan, Y. (2004). Optimism across cultures: In repsonse to the severe acute respiratory syndrome outbreak. *Asian Journal of Social Psychology, 7*, 25-34.

Johnson, D. D. P. (2004). *Overconfidence and war: The havoc and glory of positive illusions*. Cambridge, MA: Harvard University Press.

Jonsson, A.-C., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences, 34*(4), 559-574.

Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 1038-1052.

Juslin, P., Winman, A., & Hansson, P. (in press). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review*.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review, 107*(2), 384-396.

Keren, G. (1988). On the ability of monitoring non-veridical perceptions and uncertain knowledge: Some calibration studies. *Acta Psychologica, 67*(2), 95-119.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271-304.

Kirchler, E., & Maciejovsky, B. (2002). Simultaneous over- and underconfidence: Evidence from experimental asset markets. *Journal of Risk and Uncertainty, 25*(1), 65-85.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79*(3), 216-247.

Klein, C. T. F., & Helweg-Larsen, M. (2002). Perceived control and the optimistic bias: a meta-analytic review. *Psychology and Health, 17*(3), 437-446.

Kleitman, S., & Stankov, L. (in press). Self-confidence and metacognitive processes. *Learning and Individual Differences*.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*(2), 147-162.

Krizan, Z., & Windschitl, P. D. (2007). Team allegiance can lead ot both optimistic and pessimistic predictions. *Journal of Experimental Social Psychology, 43*(2), 327-333.

Krueger, J. I. (2000). The projective perception of the social world: A building block of social comparison processes. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 323-351). Dordrecht, Netherlands: Kluwer Academic Publishers.

Krueger, J. I., Acevedo, M., & Robbins, J. M. (2005). Self as sample. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 353-377). New York: Cambridge University Press.

Krueger, J. I., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology, 82*(2), 180-188.

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*(2), 221-232.

Kruger, J., & Burrus, J. (2004). Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology, 40*(3), 332-340.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.

Kruger, J., & Savitsky, K. (2006). *On the genesis of inflated (and deflated) judgments of responsibility: Egocentrism revisited*. Available at SSRN: http://ssrn.com/abstract=946245. http://ssrn.com/abstract=946245

Kruger, J., Windschitl, P. D., Burrus, J., Fessel, F., & Chambers, J. R. (in press). The rational side of egocentrism in social comparisons. *Journal of Experimental Social Psychology*.

Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review, 111*(1), 94-110.

Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). *Organizational Behavior & Human Decision Processes, 102*(1), 76-94.

Lerner, J. S., Gonzalez, R. M., Small, D. A., & Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychological Science, 14*(2), 144-150.

Liberman, V. (2004). Local and global judgments of confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 729-732.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Decision Processes, 20*(2), 159-183.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art in 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-333). Cambridge, England: Cambridge University Press.

Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. *Journal of Finance*(60), 6.

May, R. M. (1986). Inferences, subjective probability, and frequency of correct answers: A cognitive approach to the overconfidence phenomenon. In B. Brehmer, H. Jungermann, P. Lourens & A. Sevoaan (Eds.), *New directions in research on decision making* (pp. 175-189). Amsterdam: North-Holland.

McGraw, A. P., Mellers, B. A., & Ritov, I. (2004). The affective costs of overconfidence. *Journal of Behavioral Decision Making, 17*(4), 281-295.

McKelvey, R. D., & Page, T. (1990). Public and private information: An experimental study of information pooling. *Econometrica, 58*, 1321-1339.

Moore, D. A. (2004a). Myopic prediction, self-destructive secrecy, and the unexpected benefits of revealing final deadlines in negotiation. *Organizational Behavior and Human Decision Processes, 94*(2), 125-139.

Moore, D. A. (2004b). The unexpected benefits of final deadlines in negotiation. *Journal of Experimental Social Psychology, 40*(1), 121-127.

Moore, D. A. (2005). Myopic biases in strategic social prediction: Why deadlines put everyone under more pressure than everyone else. *Personality and Social Psychology Bulletin, 31*(5), 668-679.

Moore, D. A. (2007). Not so above average after all: When people believe they are worse than average and its implications for theories of bias in social comparison. *Organizational Behavior and Human Decision Processes, 102*(1), 42-58.

Moore, D. A., & Cain, D. M. (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior & Human Decision Processes, 103*, 197-213.

Moore, D. A., Kurtzberg, T. R., Fox, C. R., & Bazerman, M. H. (1999). Positive illusions and forecasting errors in mutual fund investment decisions. *Organizational Behavior and Human Decision Processes, 79*(2), 95-114.

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative social judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology, 92*(6), 972-989.

Moore, D. A., & Small, D. A. (in press). When it's rational for the majority to believe that they are better than average. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn M. Dawes*. Mahwah, NJ: Erlbaum.

Neale, M. A., & Bazerman, M. H. (1985). The effects of framing and negotiator overconfidence on bargaining behaviors and outcomes. *Academy of Management Journal, 28*(1), 34-49.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics, 122*(3), 1067-1101.

Nussbaum, S., Liberman, N., & Trope, Y. (2006). Predicting the near and distant future. *Journal of Experimental Psychology: General, 135*(2), 152-161.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *Journal of Finance, 53*(6), 1887-1934.

Odean, T. (1999). Do investors trade too much? *American Economic Review, 89*(5), 1279-1298.

Paunonen, S. V. (1989). Consensus in personality judgments: Moderating effects of targetater acquaintanceship and behavior observability. *Journal of Personality and Social Psychology, 56*(5), 823-833.

Pfeifer, P. E. (1994). Are we overconfident in the belief that probability forecasters are overconfident? *Organizational Behavior and Human Decision Processes, 58*(2), 203-213.

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

Presson, P. K., & Benassi, V. A. (1996). Illusion of control: A meta-analytic review. *Journal of Social Behavior & Personality, 11*(3), 493-510.

Pulford, B. D., & Colman, A. M. (1997). Overconfidence: Feedback and item difficulty effects. *Personality and Individual Differences, 23*(1), 125-133.

Rutter, D. R., Quine, L., & Albery, I. P. (1998). Perceptions of risk in motorcyclists: Unrealistic optimism, relative realism, and predictions of behavior. *British Journal of Psychology, 89*, 681-696.

Sanbonmatsu, D. M., Sherman, S. J., & Hamilton, D. L. (1987). Illusory correlation in the perception of individuals and groups. *Social Cognition, 5*(1), 1-25.

Santos-Pinto, L., & Sobel, J. (2005). A model of positive self-image in subjective assessments. *American Economic Review, 95*(5), 1386-1402.

Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the big five. *Journal of Research in Personality, 38*(5), 473-480.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics, 1*(1), 43-61.

Slovic, P. (Ed.). (2001). *Smoking: Risk, perception, and policy*. Thousand Oaks, CA: Sage.

Sniezek, J. A. (1990). A comparison of techniques for judgmental forecasting by groups with common information. *Group & Organization Studies, 15*(1), 5-19.

Soll, J. B. (1996). Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65*(2), 117-137.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 299-314.

Stankov, L., & Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences, 21*(6), 971-986.

Stankov, L., & Crawford, J. D. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence, 25*(2), 93-109.

Statman, M., Thorley, S., & Vorkink, K. (2006). Investor overconfidence and trading volume. *Review of Financial Studies, 19*(4), 1531.

Stone, D. N. (1994). Overconfidence in initial self-efficacy judgments: Effects on decision processes and performance. *Organizational Behavior and Human Decision Processes, 59*(3), 452-474.

Stotz, O., & von Nitzsch, R. (2005). The Perception of Control and the Level of Overconfidence: Evidence from Analyst Earnings Estimates and Price Targets. *Journal of Behavioral Finance, 6*(3), 121-128.

Svenson, O. (1981). Are we less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*, 143-151.

Teigen, K. H., & Jorgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology, 19*(4), 455-475.

Thompson, S. C., Armstrong, W., & Thomas, C. (1998). Illusions of control, underestimations, and accuracy: A control heuristic explanation. *Psychological Bulletin, 123*(2), 143-161.

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87*(3), 506-516.

Viscusi, W. K. (1990). Do smokers underestimate risks? *Journal of Political Economy, 98*(6), 1253-1269.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*(5), 806-820.

Weinstein, N. D. (1998). Accuracy of smokers' risk perceptions. *Annals of Behavioral Medicine, 20*(2), 135-140.

Windschitl, P. D., Kruger, J., & Simms, E. (2003). The influence of egocentrism and focalism on people's optimism in competitions: When what affects us equally affects me more. *Journal of Personality and Social Psychology, 85*(3), 389-408.

Windschitl, P. D., & Wells, G. L. (1997). Behavioral consensus information affects people's inferences about population traits. *Personality and Social Psychology Bulletin, 23*(2), 148-156.

Winman, A., Hansson, P., & Juslin, P. (2004). Subjective probability intervals: How to reduce overconfidence by interval evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1167-1175.

Woloshin, S., Schwartz, L. M., Black, W. C., & Welch, H. G. (1999). Women's perceptions of breast cancer risk: How you ask matters. *Medical Decision Making, 19*(3), 221-229.

Footnotes

[1] While it is possible for the majority to be either above or below average (in skewed distributions), it is statistically impossible for the majority to be above (or below) the median.

[2] Assuming that L_i has a non-zero mean would not substantially alter the conclusions of the theory. The theory would simply predict that a person with prior overestimation, for example, would exhibit less overestimation (or, possibly, underestimation) after an easy quiz and even more overestimation after a difficult quiz.

[3] Having v_L differ between i and j will not qualitatively affect the results in any way.

[4] For a derivation, see, Berger (1980). This formula holds for other distributions besides the normal distribution (see Diaconis & Ylvisaker, 1979). Chambers and Healy (2007) investigate conditions under which the posterior expectation is guaranteed to lie between the prior mean and the realization of the signal.

[5] Note that we are focusing on the association between precision in beliefs and the accuracy (with respect to estimation and placement) of those same beliefs. Elsewhere we examine the separate question of how precision in prior expectations should affect overestimation and overplacement in subsequent, updated beliefs (Healy & Moore, 2007).

[6] To run through the experimental stimuli yourself, visit http://cbdr.cmu.edu/roe and log in using the participant ID number 0000.

[7] The questions from the quizzes and the data are available at http://www.andrew.cmu.edu/user/donm/CRO

[8] Note that participants could have guaranteed their ability to predict their own score perfectly by predicting scores of zero for self and intentionally getting all the quiz questions wrong. While this would be worth $4 in earnings, it is not a wise strategy because $4 represents

a small fraction of what a higher performance was worth. By doing well on the quiz, participants could earn up to $25, and should have expected to earn $12.50, on average, for doing well. Such sandbagging would be most obvious on the easy quizzes. In fact, of the 492 easy quizzes, we observe only eleven scores of zero or one from nine different participants. Only three of these followed ex ante predictions of scores below two. We conclude that sandbagging did not have a large influence on our results.

[9] Answer: "The Eagles: Their Greatest Hits"

[10] This figure is above 90% due to the 20% of estimates on which people reported that they were 100% sure they knew the score exactly.

Table 1

Participants' overestimation of their own performances, measured at the interim phase, over the six trial blocks for the three different quiz difficulties. (Standard deviations in parentheses.)

	Block Number						
	1	2	3	4	5	6	Overall
Easy	-0.40 (1.07)	-0.20 (0.79)	-0.29 (0.83)	-0.10 (0.78)	-0.10 (0.82)	-0.22 (1.20)	-0.22 (0.93)
Medium	-0.13 (1.65)	0.01 (1.14)	0.05 (1.25)	-0.05 (1.16)	-0.15 (1.33)	0.31 (0.94)	0.01 (1.27)
Hard	1.15 (1.63)	0.69 (1.62)	0.87 (1.61)	0.71 (1.22)	0.69 (1.37)	0.63 (1.49)	0.79 (1.50)

Table 2

Participants' overplacement of their own performances, measured at the interim phase, over the six trial blocks for the three different quiz difficulties. (Standard deviations in parentheses.)

	Block Number						
	1	2	3	4	5	6	Overall
Easy	0.56 (2.70)	0.55 (2.45)	0.08 (2.84)	0.59 (2.13)	0.75 (2.44)	0.36 (2.89)	0.48 (2.59)
Medium	-0.25 (3.82)	-0.23 (4.14)	-0.10 (4.03)	0.41 (3.46)	0.22 (3.99)	0.15 (4.10)	0.04 (3.91)
Hard	-1.46 (2.54)	-1.47 (2.45)	-1.52 (2.51)	-1.19 (2.19)	-1.10 (2.17)	-1.39 (2.51)	-1.36 (2.39)

Table 3

Regressions predicting interim score estimates for self and RSPP. Quiz ease is quantified as the mean score among all participants on that particular quiz. (Standard errors in parentheses.)

Model 1 predicting interim beliefs about own performance		Model 2 predicting interim beliefs about RSPP's performance	
Independent variable	Unstandardized *B* coefficient	Independent variable	Unstandardized *B* coefficient
Prior estimated score for self	.08*** (.02)	Prior estimated score for RSPP	.28*** (.03)
Own actual score	.86*** (.02)	Own actual score	.31*** (.02)
Quiz ease (mean score on quiz)	.02 (.02)	Quiz ease (mean score on quiz)	.35*** (.02)
R^2	.90***	R^2	.67***

*** $p < .001$

Figure 1. An example of the theory's prediction of beliefs about performance by self and others on a ten-item trivia quiz as a function of the actual score of the person doing the predicting, assuming the person expected a score of five prior to taking the quiz.

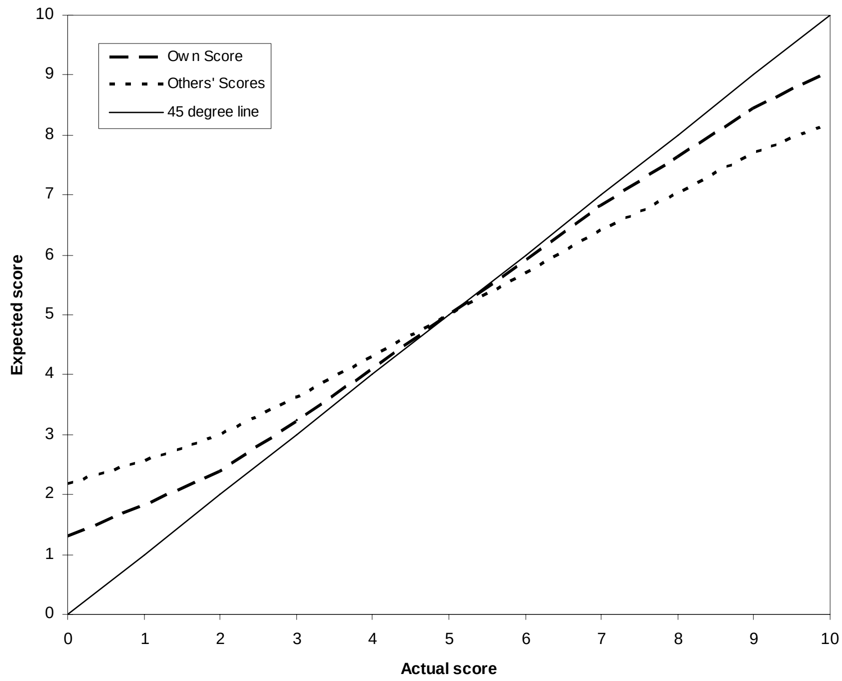

Figure 2. Participants' estimated scores for self and other as a function of quiz difficulty. Error bars show standard errors.