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 May 22, 2007

Don A. Moore: Tepper School of Business, Carnegie Mellon University. Paul J. Healy: 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 Peter Bryant, 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 the 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 believe that they are worse than others; on easy tasks, people underestimate their actual performances but 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

Researchers have offered overconfidence as an explanation for persistent high rates of entrepreneurial entry, despite the frequency of entrepreneurial failure (Camerer & Lovallo, 1999). Malmendier and Tate (2005) used overconfidence to explain the high rates of corporate merger and acquisition, despite the fact that they so often fail. Howard (1983) and Johnson (2004) implicated overconfidence in the causes of war. Odean (1998) argued that overconfidence could explain the excessively high rate of trading in the stock market, despite the costs of trading (Odean, 1999). Neale and Bazerman (1985) blamed overconfidence as a cause for both labor strikes and litigation. In the words of one popular text on judgment, which suggests that overconfidence contributed to the nuclear accident at Chernobyl and to the explosion of the Space Shuttle Challenger, “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.

How has Overconfidence Been Studied?

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. Roughly 64% of empirical papers on overconfidence examine overestimation. Some findings from this literature include the following: students overestimate their performance on exams (Clayson, 2005), physicians overestimate the accuracy of their diagnoses (Christensen-Szalanski & Bushyhead, 1981), people overestimate how much control they have (Presson & Benassi, 1996), and people overestimate the speed with which they can get work done (Buehler, Griffin, & Ross, 1994).

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. Roughly 5% of empirical overconfidence papers (12 of them) examine this phenomenon. Following Larrick, Burson, and Soll (2007), we will call this overplacement. Sample findings from this literature include the following. In one study, 37% of one firm's professional engineers placed themselves among the top 5% of performers at the firm (Zenger, 1992). In a survey of high school seniors, 25% rated themselves in the top 1% in their ability to get along with others (College Board, 1976-1977). But perhaps the most frequently cited result from this literature is that 93% of a sample of American drivers and 69% of a sample of Swedish drivers reported that they were more skillful than the median driver in their own country (Svenson, 1981).[1] Note that research on overplacement 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).

The third way overconfidence has been measured 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

---

[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.

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 knew 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).

Researchers routinely assume, either explicitly or implicitly, that the different types of overconfidence are the same (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. However, Kwan and colleagues (2004) have demonstrated the importance of distinguishing overestimation and overplacement from each other, and how this sort of clarification can resolve longstanding arguments regarding the phenomenon in question that arise from inconsistent research findings. In this paper we ask whether the three varieties of overconfidence we have identified are merely differing operationalizations of the same underlying construct. If not, how are they related to each other?

To answer these questions, we first review the literature on overconfidence and identify empirical inconsistencies and methodological problems. 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 key predictions of our new theory.

The Trouble with Overconfidence

There are three notable problems with research on overconfidence.

Problem 1: The confounding of overestimation and overprecision

The first problem with overconfidence research is that the most popular research paradigms confound overestimation and overprecision, making it impossible to distinguish the relative influence of each. These research paradigms measure people's confidence of 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. For instance, when participants estimated at least a 99% chance they had gotten a question correct, they were actually correct only 87% of the time, on average. 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 to distinguish overprecision from overconfidence, respondents must be able to express excessive certainty in an underestimate of their performance. 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 solution to this problem is measuring perceptions of performance over a set of items. It is easy to see how measuring beliefs about performance over a set of items can reduce the confound between overprecision and overestimation. It is possible, for instance, for someone to underestimate how well he has done, yet be excessively confident that his low estimate is correct. Estimating performance over a set of items usually entails making a frequentistic judgment (such as the number of items correct), whereas estimation of performance at the item level is more often elicited as a probabilistic judgment. There is a vigorous debate over frequentistic vs. probabilistic judgment. A number of researchers have found that overconfidence is reduced or eliminated when people are given frequentistic judgment tasks, such as estimating performance across a set of items (Gigerenzer, 1993; Gigerenzer & Hoffrage, 1995; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Juslin, Olsson, & Björkman, 1997; Sniezek & Buckley, 1991). While other researchers argue that frequentistic judgments are not necessarily more accurate than probabilistic judgments (Griffin & Buehler, 1999), there is consensus that frequency judgments across a set of items are less prone to overconfidence than are judgments of correctness at the item-level.

Some have suggested that the difference between frequentistic and probabilistic judgments may be due to the human mind being better adapted to reason frequentistically (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). While that may be so, another possibility is simply that overconfidence in item-confidence judgments is attributable to overprecision. Item-level (probabilistic) confidence judgments will be biased upward by both overestimation and overprecision, whereas set-level (frequentistic) judgments will not necessarily be biased upward by overprecision.

The fact that most of the evidence for the existence of overconfidence confounds overestimation and overprecision raises important questions about which effect is responsible for the pervasiveness of apparent overconfidence in the results. It is possible that both causes share equally in the observed results, but it may be that one alone is primarily responsible. Prior data from item-confidence judgments cannot answer this question. We must ask it using a different paradigm.

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 (Fu, Koutstaal, Fu, Poon, & Cleare, 2005; Griffin & Tversky, 1992; Kirchler & Maciejovsky, 2002; Koriat, Sheffer, & Ma'ayan, 2002; Lichtenstein & Fischhoff, 1977; Lichtenstein, Fischhoff, & Phillips, 1982; Sniezek, 1990; Stankov & Crawford, 1997). It is those individuals with the best performances on any given task who are most likely to underestimate their actual performances (Klayman et al., 1999; Krueger & Mueller, 2002; Kruger & Dunning, 1999; Larrick et al., 2007).

Erev, Wallsten, and Budescu (1994) offered what has been called a Thurstonian explanation for the fact that overestimation occurs on hard tasks but underestimation occurs for easy tasks. Their theory, following in the tradition of Thurstone (1927), takes into account the fact that any judgment is likely to include some error component. When people have imperfect knowledge of their own performances, the error in their estimates will make those estimates regressive (Burson, Larrick, & Klayman, 2006; Krueger & Mueller, 2002). It is, after all, easier to underestimate than overestimate your score on a test when you get everything right. As a result, people underestimate their performance when it is high.

Underestimation has also been documented in other domains other than beliefs about one's own prior performance, including the illusion of control, the planning fallacy, and optimism about future events.

The illusion of control. Research on the illusion of control has generally shown that, when people have no control over some event, they frequently act as if they have some sort of control (for reviews see Presson & Benassi, 1996; Thompson, Armstrong, & Thomas, 1998). While these results make it appear that people overestimate their control, the research paradigms have generally focused on situations in which people have very little control. Just as overestimation of performance is most likely to occur when performance is low, overestimation of control is most likely to occur when control is low. It is, after all, difficult to underestimate one's control if one has none. Studies find that people actually tend to underestimate their control when it is high (Alloy & Abramson, 1979; Jenkins & Ward, 1965).

The planning fallacy. The planning fallacy documents the tendency for people to underestimate how long it takes to get things done, or overestimate their own productivity (Buehler et al., 1994). It is, of course, easiest to underestimate how long it will take to get something done when the task will be time-consuming. When the task doesn't take long to complete (i.e., it's easy), people are more likely to overestimate completion times—thereby underestimating their future performance (Boltz, Kupperman, & Dunne, 1998; Burt & Kemp, 1994).

Pessimism about the future. While people prefer to imagine bright futures for themselves (Markus & Nurius, 1986), there is clear evidence showing pessimism about future events under some circumstances. For instance, shortly after September 11th, 2001, a sample of Americans estimated they had a 20% chance of being injured in a terrorist attack in the coming year (Lerner, Gonzalez, Small, & Fischhoff, 2003). This is almost certainly a radical overerstimate, because there were no terrorist attacks that resulted in injuries in the domestic United States in the year 2002. In another study, smokers estimated their risk of dying of lung cancer at about 33%, when in fact the actual base rate among smokers is below 10% (Viscusi, 1990). Women estimate their chances of falling ill with breast cancer by as much as eight times the actual risk (Woloshin, Schwartz, Black, & Welch, 1999). Chinese people overestimated their actual chances of falling ill with Severe Acute Respiratory Syndrome (SARS) during the disease's outbreak in China (Ji, Zhang, Usborne, & Guan, 2004). People overestimate their risk of dying in the coming year (Fischhoff, Bruine de Bruin, Parker, Milstein, & Halpern-Felsher, 2006). It is notable that all these overestimates of risk occur in domains where the outcome has a low probability. All of these are also “easy” tasks, in the sense that the positive outcomes (*not* being the victim of a terrorist attack or *not* contracting cancer) is by far the most likely outcome. Again, when success is likely, people tend to underestimate it.

Underplacement

It is also common to find situations in which people underplace their performances relative to others. On hard tasks, such as juggling or unicycle riding, people routinely rate themselves, on average, below average (Kruger, 1999). But difficulty can also arise from other sorts of challenges; people believe they would not be able to handle the death of a loved one as well as others would (Blanton, Axsom, McClive, & Price, 2001).

Difficult competitions. 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. 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. The most frequently cited evidence of comparative optimism are Weinstein’s (1980) findings—the belief that one is more likely to experience positive events than are others, and that one is less likely to experience negative events than are others. However, it just so happened that Weinstein asked his participants about positive common events (such as owning your own house) and rare negative events (such as attempting suicide). When event commonness and valence are unconfounded, results reveal that while both influence judgments of comparative likelihood, the effect of event commonness is roughly four times the size of valence (Kruger & Burrus, 2004). That is, people believe that common events (even undesirable ones) are more likely to happen to them than to others and they believe that rare events (even desirable ones) are less likely to happen to them than to others (J. R. Chambers, Windschitl, & Suls, 2003).

Underprecision

Results showing underprecision are rarer than results showing underestimation or underplacement. Since Alpert and Raiffa's (1969/1982) original demonstration, numerous studies have examined the tendency for people's judgments to reflect excessive confidence that their beliefs are true and accurate. The most common paradigm used in this literature is to have people specify confidence intervals around their estimates to indicate the precision of their beliefs. Research typically finds that 90% confidence intervals contain the true answer somewhere between 30% and 50% of the time (Teigen & Jorgensen, 2005). Soll and Klayman (2004) note that the standard paradigm tends to overstate the true size of the overprecision effect, but that they find that overprecision is nevertheless robust. Some research has found that 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; Winman, Hansson, & Juslin, 2004), and overprecision tends to be stronger for unfamiliar tasks (Block & Harper, 1991).

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 (Moore & Small, in press-a). 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).

These inconsistencies have produced some serious disputes. For instance, Viscusi (1990) asked smokers to estimate their chances of contracting lung cancer and found that they overestimated this small risk. On the other hand, when Weinstein (1998) asked smokers whether they were more or less likely than other smokers to get lung cancer, they told him they were less likely than others to fall ill. Slovic (2000) and Viscusi (2000) have argued bitterly on the question of whether smokers are overconfident or not.

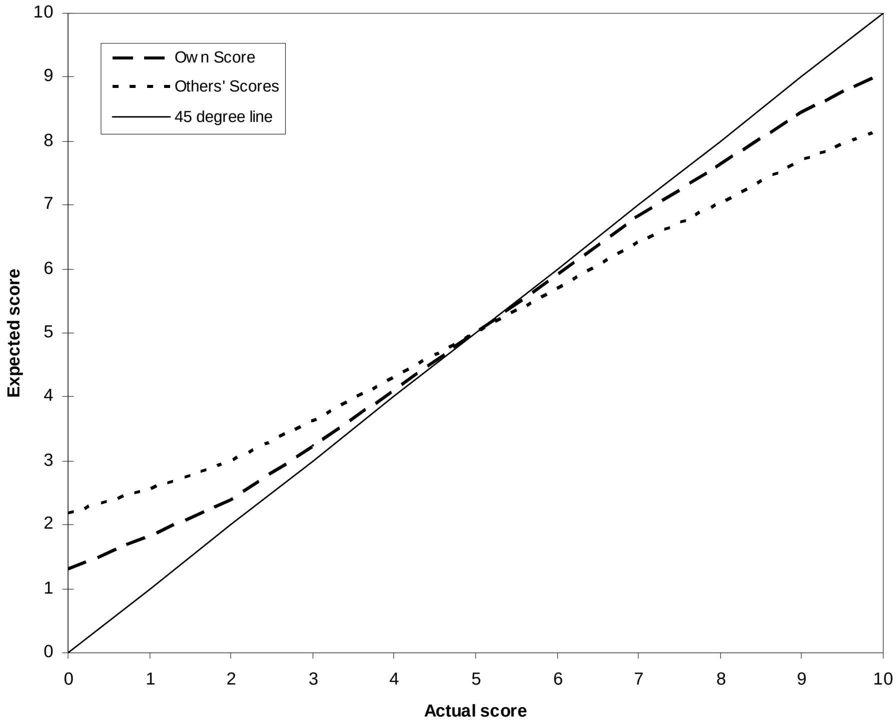
In this paper, we attempt to address the three problems identified above, both theoretically and empirically. Our work builds on that of Larrick, Burson, and Soll (2007), who present an insightful analysis of the statistical relationships between overestimation, overplacement, and actual performance. Our theory goes beyond theirs by considering the underlying beliefs regarding own and others' performance that produce these relationships, and by considering overprecision. 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. 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.

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.



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*, believes her score, denoted by X_i, to be the sum of 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 subject's 'luck' or skill. Thus, $X_i = S + L_i$.

Before seeing the contents of the quiz, the individual has some 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 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 and others' scores are realizations of 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 (her own or others'), after taking the quiz but before learning her own score or the score of anyone, after learning her own score, but not the scores of others, and finally, after learning the scores of others.

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 his performance but won't know his 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

---

[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 affect the results in any way.

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 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

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

which 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

---

[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.

score ($x_i < E[X_i|y_i] < m_S$), implying overestimation after hard quizzes.[5]

Note that these conclusions 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

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

---

[5] Note that these arguments are taken on average because we can rarely observe the signal y_i in practice. 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.

which 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 5 out of 10 on the quiz.

Overprecision

Our theory has little to say about the origins of overprecision. However, if the theory is correct regarding the origins of overestimation and overplacement, then we should expect precision of beliefs to be correlated with each of them. Specifically, precision in beliefs ought to be associated with reduced bias in estimation of one's self.[6] Furthermore, precision in beliefs about others ought to reduce bias in placement of one's self relative to others.

What is the logic underlying these predictions? Our theory holds that regressiveness in estimates of self leads to over- and underestimation, and that regressiveness in estimates of others lead to over- and underplacement. Regressiveness, for its part, arises directly from imperfection of information. 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 who are unsure of themselves 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.

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.

---

[6] Note that we are focusing on the effect of precision in beliefs on 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).

Sometimes, people appear to neglect priors (such as base rates) and overweight new evidence (Grether, 1990). Under other circumstances, people appear to overweight priors and neglect new evidence (Edwards, 1968; McKelvey & Page, 1990). Which of these errors people commit depends on the order and form in which they acquire information (Griffin & Tversky, 1992; Hogarth & Einhorn, 1992; Wells, 1992).

The important thing, however, is that while people are not perfect Bayesians, their judgments adhere to the basic logic, if not the precise formulas, of Bayes’s Law. Their posterior judgments generally fall between their priors and the new evidence, and judgments sometimes adhere to Bayesian principles remarkably well (Ajzen, 1977; Griffiths & Tenenbaum, 2006; Kersten, Mamassian, & Yuille, 2004). 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.

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. Prior research has not done this. Our second contribution is the application of the Bayesian principle of updating from prior beliefs based on data. 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). Like the Thurstonian account for overconfidence, our theory rests on the assumption that people make imperfect estimates of their own performances. Like Thurstonian theories, ours predicts that on any given task, it is those who perform worst who are most likely to overestimate their performances. However, ours is not a straight Thurstonian account, as the random error present in human judgment is not sufficient to produce systematically biased estimates of performance when that performance is not constrained by floor or ceiling effects. Thurstonian accounts of overconfidence have focused on probabilistic item-confidence judgments, which are bounded by the probability scale.

Our theory, by contrast, would predict that estimates of performance tend to regress toward people's prior beliefs, and that this holds even when the range on which performance is measured can range from infinitely good to infinitely bad. Empirical results show that the hard/easy effect persists, even for tasks on which neither performance nor its measurement is bounded by floors or ceilings (Moore & Small, in press-b). Our theory can account for these results better than can Thurstonian theories. The central role our theory gives to prior expectations and their updating makes it more of a Bayesian than a Thurstonian theory. 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. These explanations point out that research paradigms examining overestimation have long employed particularly difficult (even "tricky") trivia questions selected by the experimenter without regard to the issue of whether these items were in any way representative of the problems people must solve in everyday life. When the problems people are asked to solve are sampled more broadly, researchers find that overestimation of performance tends to decrease (Gigerenzer et al., 1991; Juslin et al., 1997).

Brunswikian explanations for overconfidence hold that overconfidence is an artifact of the way items were selected. In particular, that experimenters chose "tricky" items that fooled participants into thinking they knew the answer when in fact they did not. 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. For instance, Gigerenzer and colleagues (1991) randomly picked 25 German cities with populations above 100,000 and presented their participants with all 300 pairwise comparisons between them. Participants had to judge which of two cities had the larger population. They found that their participants were not overconfident regarding the correctness of their answers on these questions.

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. Most Germans will correctly identify Berlin as more populous than Würzburg. These simple items are more likely to produce underestimation than overestimation, since it turns out to be difficult to overestimate your correctness when you are 100% correct. Consistent with this explanation, Soll (1996) found that item difficulty can account for much of the reduction in overconfidence on representative sets. 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) was built on the Brunswikian or Lewinian principle that "to understand human behavior, one first has to analyze the texture of the environment" (p. 193). Fiedler's BIAS model focuses on the differential information people obtain from their environments regarding themselves and their ingroups rather than other people and outgroups. Given the key role that differential self/other information plays in our theory, Fiedler's (1996) BIAS model is clearly related. Fiedler (1996; 2000b) proposed that better information about the self than others could explain overplacement effects, as well as a variety of other important psychological phenomena. Like ours, Fiedler's model begins with the fact that people usually have more information about themselves than they do about others, because they observe more instances of their own behavior than that of others. Since desirable behaviors such as being cooperative and friendly are generally more frequent than undesirable behaviors such as being rude and phony, people will conclude that positive traits are more descriptive of themselves than of others (e.g., Alicke, 1985).

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 or for rare behaviors. Second, our theory elaborates on Fielder's model by explicitly allowing for the possibility that people use information about themselves to make inferences about others (Moore

& Small, in press-a). 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 theory's predictions 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, and for generalizing the explanation to circumstances in which signals (and beliefs) are not constrained by ceilings or floors.

Integrative theories

Several researchers have proposed interesting and useful theories that help integrate and reconcile the Thurstonian and Brunswikian explanations for overconfidence (Dougherty, 2001; Juslin et al., 1997; Soll, 1996). These integrative theories show how the two perspectives are, in many ways, consistent with one another. However, because all of these theories focus on explaining results from the item-confidence paradigm, they are of more limited value outside this specific paradigm.

Hypotheses

Our theory makes four basic predictions regarding overestimation and overplacement:

Overestimation: When people perform worse than they expected, they will overestimate their performances on average.

Underestimation: When people perform better than they expected, they will underestimate their performances on average.

Overplacment: When people perform better than they expected, they will, on average, overplace their performance relative to the performance of others.

Underplacement: When people perform worse than they expected, they will, on average, underplace their performance relative to the performance of others.

We expect all four of these effects to follow straightforwardly from updating prior beliefs using imperfect information about performance. In our experimental paradigm, as in most of life, people have imperfect information about the performance of both self and others, but they have much better information about themselves than they do about others.

We are agnostic with respect to the origins of overprecision, as our theory does not have much to say about how precise people's judgments ought to be. None of the hypotheses described above depend on the accuracy in people's estimations of performance. However, 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 hard/easy effect will be associated with greater precision in estimates of one's own performance.

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. In addition, we ask whether there are reliable differences between people with respect to how much they display each type of overconfidence.

ILLUSTRATIVE STUDY

In addition to investigating the consistency between different varieties of overconfidence, we are interested in asking whether any of them correlate with measures of individual differences between people (Stanovich & West, 1998). Of course, the first question we must ask of our data in this regard is whether we find stable individual differences in measures of overconfidence. In other words, do our repeated measures of each variety of overconfidence represent reliable measures?

If the answer to this question is affirmative, then we can ask whether individual differences between people can account for the differences in overconfidence we observe. There have been some attempts to examine individual differences in the propensity to be overconfident. While not every study that looks has found a relationship (Jonsson & Allwood, 2003), there are some that claim to have identified the traits that predict overconfidence. We included some measures because prior research has found them to be related to overconfidence: the “big five” dimensions of personality (Schaefer, Williams, Goodie, & Campbell, 2004), gender (Niederle & Vesterlund, in press; Pulford & Colman, 1997; Soll & Klayman, 2004), narcissism (Ames & Kammrath, 2004; Campbell et al., 2004), and cognitive abilities (Kleitman & Stankov, in press; Stankov & Crawford, 1996). In addition, because we expected them to be related to self-enhancement, we measured overclaiming (Paulhus, Harms, Bruce, & Lysy, 2003), self-esteem (Rosenberg, 1965), and the generalized sense of power (Anderson, John, & Keltner, 2005).

Because we suspected that they might be related to individuals' willingness and ability to understand others, we included measures of empathy and perspective-taking (Davis, 1983), and social comparison orientation (Gibbons & Buunk, 1999). Finally, because we suspected that it might be related to confidence in one's judgments and willingness to endure uncertainty we included measures of political conservatism (Jost, Glaser, Kruglanski, & Sulloway, 2003).

Method

Participants

Participants were 82 students from a private research university in the United States. 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*." The entire experiment was run on computers via a web server, but all participants were present in the experimental laboratory.[7] 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

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.[8] Each topic included one easy, one hard, and one medium difficulty quiz. Each participant encountered a different ordering of these 18 quizzes. The ordering was random, subject to the

---

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

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

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 11 bars to indicate probability. When they slid one bar, the other 10 automatically adjusted proportionally so that the total probability across all 11 bars was always equal to 100%. These 11 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). These are their *ex ante* predictions, made 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 scores[9] and the RSPP's score again. Then they were shown the correct answers to the quiz and were asked to grade themselves. They were then again asked to 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.

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 predicted scores using the eleven probability estimates by multiplying each of the eleven possible scores by its reported probability, and summing these.

Overestimation. A participant's actual score was subtracted from his or her reported expected 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

---

[9] 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, in expectation, 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 is not a large influence on our results.

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. Prior research has generally measured overprecision by having participants specify confidence interval ranges. Unfortunately, Soll and Klayman (2004) have shown that this method has the problem that it tends to overstate the size of the overprecision effect. Our paradigm allows us to obtain better measures of beliefs than most studies do, because we can measure perceived probability of every possible outcome. One of the other 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, Winman, & Hansson, in press). We test for overprecision in these data in two ways. First, we compare the shape of this subjective belief distribution with the actual distribution in scores. Second, we compare rated probabilities with actual probabilities for the scores participants actually obtained.

Individual difference measures

After all 18 rounds were done participants completed 17 measures of individual differences. The 17 measures appeared in a different randomly determined order for each participant. Those 17 measures were as follows: the "big five" dimensions of personality (John & Strivastava, 2000), gender, narcissism (Raskin & Terry, 1988), cognitive reflectiveness (Frederick, 2005), overclaiming (Paulhus et al., 2003), self-esteem (Rosenberg, 1965), the generalized sense of power (Anderson et al., 2005), the empathy and perspective-taking subscales of the Interpersonal Reactivity Index (Davis, 1983), social comparison orientation (Gibbons & Buunk, 1999), and a basic measure of political conservatism. This last measure included two questions: "*How liberal or conservative do you tend to be when it comes to social issues?*" and "*How liberal or conservative do you tend to be when it comes to economic issues?*" Participants were provided with a 7-point scale on which to respond, with endpoints labeled, "*very liberal*" and "*very conservative*."

Payoffs

After participants had completed these questionnaires, 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 ($M = .69$).

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 overestimation.

Note that in the analyses below, 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

Overall, our participants do display a modest degree of overestimation. At the interim phase, they report believing that they scored 0.2 (*SD* = 1.31) points higher than they did, and this overestimate is significantly greater than zero, t (1475) = 5.82, $p < .001$. However, this tendency is powerfully moderated by quiz difficulty.

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, $p = .003$, $\eta^2 = .03$. See Table 1.

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)

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 exhibit underplacement at the interim phase (*M* = -.27), t (1475) = -3.35, *p* = .001. And their posterior beliefs continue to display underplacement (*M* = -.37), t (1475) = -4.69, *p* < .001. This underplacement is, as expected, moderated by task difficulty.

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, η^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.

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.)

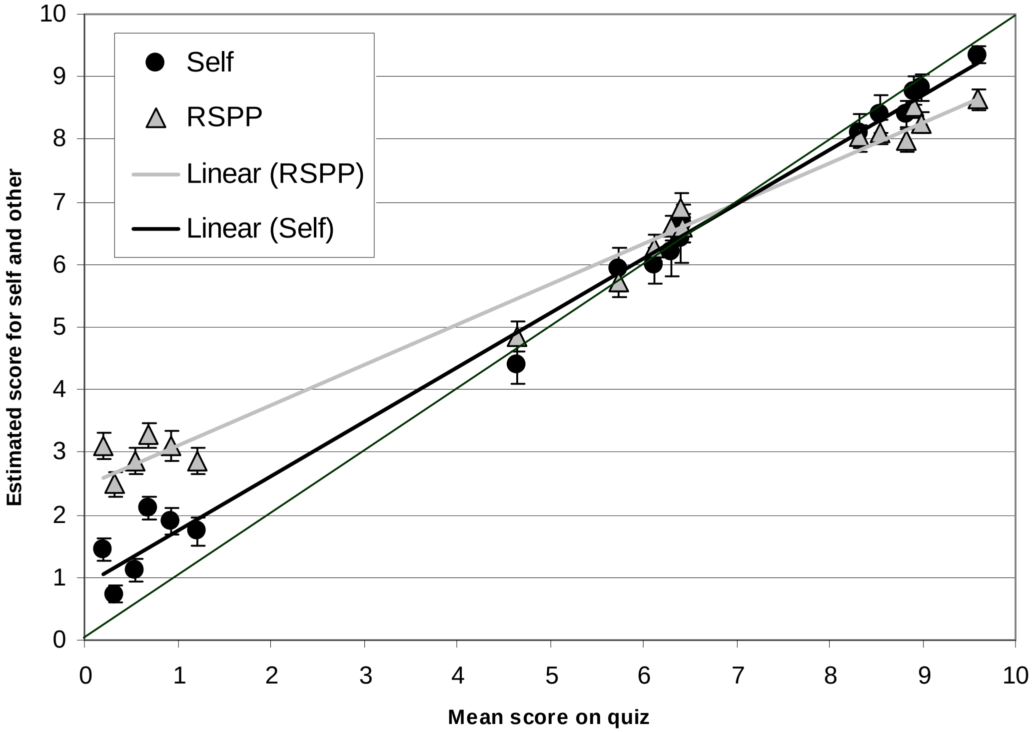
	Block Number						
	1	2	3	4	5	6	Overall
Easy	0.56 (2.70)	0.55 (2.45)	0.08 (2.84)	0.59 (2.13)	0.75 (2.44)	0.36 (2.89)	0.48 (2.59)
Medium	-0.25 (3.82)	-0.23 (4.14)	-0.10 (4.03)	0.41 (3.46)	0.22 (3.99)	0.15 (4.10)	0.04 (3.91)
Hard	-1.46 (2.54)	-1.47 (2.45)	-1.52 (2.51)	-1.19 (2.19)	-1.10 (2.17)	-1.39 (2.51)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

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 easy affects overestimation and overplacment in opposite ways. See Figure 2.

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



Two features of Figure 4 deserve mention. First, the regressiveness of estimates for both self and the RSPP are 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'?"[10]

---

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

Overprecision

Unlike overestimation and overplacement, participants' judgments *do* appear to show systematic overprecision at the prior stage. The entire sample of 1476 quiz scores has variance of 16.97. But participants' prior judgments have a smaller variance, suggesting they were inappropriately sure they knew what their scores would be. The variance of participants' estimates of themselves is 6.37 and of the RSPP is 6.51, both of which are significantly smaller than 16.97, *p*s < .001. But that could just be because the distribution of actual quizzes was more strongly bimodal than they were expecting, as the quizzes included so many surprisingly easy and surprisingly difficult quizzes, which pushed up the variance on actual quiz scores.

More interestingly, their estimates of the RSPP show lower variance (and imply greater precision) than the actual set of scores on each specific quiz, both at the interim and the posterior phase. Across the 18 quizzes, the average variance in scores is 5.19. However, at the interim phase, the variance in the average participant's estimate of the RSPP's score is only 2.86 and this shrinks to 2.67 at the posterior phase. Both of these represent significant underestimation of the actual variance, *p*s < .001. Note that it does not make sense to perform this analysis on participants' estimates of their own scores because they observe useful private signals that ought to produce more precise estimates of their own scores.

The tendency toward overprecision in estimation of the RSPP's score is moderated by test difficulty. We computed an index of interim overprecision[11] by subtracting the variance of each participant's estimate of the RSPP from the actual variance on that quiz. When this

---

[11] Note that we observe the same results if we compare the variance of posterior (rather than interim) estimations of the RSPP with actual quiz variances. We think these analyses are less interesting, however, if performed on prior estimates. Also, the comparison of variances between beliefs and the actual distribution of scores does not make sense for the self at either the interim or posterior phase, because then people have excellent knowledge about their own particular quiz performances, and so the variance in each person's estimate ought to be smaller than the variance across all participants.

measure is subject to a 6 X 3 within-subjects ANOVA, the result reveals a significant main effect for difficulty, F (2, 162) = 1647, $p < .001$, $\eta^2 = .95$. Participants' judgments reveal greater overprecision for medium quizzes (*mean variance* = 6.63) than for easy quizzes (*mean variance* = 2.17), and underprecision for difficult quizzes (*mean variance* = -1.78). Neither the main effect for block, F (5, 405) = 1.56, $p = .17$, nor its interaction with difficulty, F (10, 810) < 1, $p = .96$, attains significance.

The reason for the significant difference across quiz difficulties is that the precision implied by participants' estimates of performance is not sufficiently responsive to changes in the variance of actual quiz scores. Because the variance in medium quizzes is quite large (*mean variance* = 9.87), participants' estimates (*mean variance* = 3.12) appear too precise. Because variance in the difficult quizzes is so small (*mean variance* = 1.48), participants' estimates (*mean variance* = 3.05) appear insufficiently precise.

Our theory offers a hint regarding why we might be most likely to observe underprecision in participants' estimates of others' scores on difficult quizzes. We have argued that overestimation occurs on difficult tests in part because when performance is low, it is easier to overestimate it than undererstimate it. Just so, when variance on a quiz is exceptionally low, it is easier to overestimate it than to underestimate it. As with estimations of performance, the precision of those estimates is insufficiently sensitive to reality (see Budescu & Du, in press).

An additional measure of overprecision. We should admit that the comparison of variances is an incomplete measure of overprecision, because it fails to take actual accuracy into account. Not all precision is overprecision. There is, in fact, a positive correlation between precision and accuracy. The smaller the variance in people's interim estimates of their own scores, the smaller the error in their estimates ($r = .55$, $p < .001$). The same is true for people's estimates of the RSPP: the smaller the variance in their interim estimates, the smaller the error in those estimates ($r = .25$, $p < .001$). Consider, for instance, a quiz on which no one gets anything right. An individual who reports that he is sure (100% confidence) that the RSPP has score a 0 would not be overestimating the precision of his beliefs. An alternative approach is to examine the one score an individual estimates is most likely. When takers of a difficult quiz report that a score of 2 is most likely, how often did they, in fact, get just 2 correct?

At the interim phase, participants report being, on average, 73% confident that they know their own scores but are correct only 59% of the time, t (1476) = 18.84, $p < .001$. They report being 43% sure they knew what the RSPP had scored, but are only correct 31% of the time, t (1476) = 17.86, $p < .001$. When averaged across all rounds and all belief elicitations for both self and the RSPP, participants reported, on average, being 44% confident that they had correctly identified the score. In actuality, they were only correct 29% of the time, and participants' 44% estimate is significantly greater than this figure, t (81) = 8.64, $p < .001$. Overprecision measured this way is greater than zero for all five belief elicitations (all ps < .001), all quizzes (all ps < .005), all levels of difficulty (all ps < .001), and all blocks (all ps < .001).

The relationship between overprecision, overestimation, and overplacement. As expected, 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 variance (i.e., lack of precision) and *under*estimation is positive ($r = .45$, $p < .001$). Participants' tendency to underestimate their scores on easy quizzes was smaller among those who reported narrow confidence intervals. On the hard quizzes, the correlation between variance and *over*estimation is also positive ($r = .46$, $p < .001$), because the more precise the estimate, the less the overestimation. The tendency to overestimate one's score on the hard quiz was associated with narrower confidence intervals.

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 variance (i.e., lack of precision) and *over*placement is positive ($r = .34$, $p < .001$). The broader the confidence interval in estimating others, the more the overplacement. On hard tests, the correlation between variance and *under*placement is also positive ($r = .26$, $p < .001$) because the broader the confidence interval in estimating the RSPP, the less the overplacement. In other words, precise estimates decrease the tendency to underplace one's own performance on hard quizzes.

Individual differences

We would like to ask whether the measures we took of individual differences between our participants are correlated with any of our overconfidence measures. But first, it makes sense to ask whether any of the three types of overconfidence are consistent within-individual. If measures of overconfidence do not vary consistently between people, then the search for individual-difference correlates of overconfidence is likely to prove fruitless. We have multiple measures of overconfidence for each person. What is the test-retest reliability of these different measures?

The alpha reliability for the 18 measures of overestimation is .21. Clearly this falls short of satisfactory reliability. The situation is much the same with overplacement: its alpha reliability is only .29. Overprecision is different, however. The average overprecision measure yields an alpha reliability of .95.

Correlations between our measures of overconfidence and a variety of individual difference measures appear in Appendix A. While three correlations in this table are significant at the .05 level, that criterion is not sufficiently stringent given that we're casting about for possible associations between our 17 individual difference variables and our three overconfidence measures. After the Bonferroni adjustment to our significance criterion for these 51 tests, none of the correlations is significant.

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.

Table 3

Regressions predicting post-test 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	.35*** (.02)
R^2	.90***	R^2	.67***

*** $p < .001$

These process results do not distinguish our theory from any alternative theoretical explanation for the negative correlation between overestimation and overplacement. We know of no alternative explanation that can account for the set of results we present. Nevertheless, we believe it is valuable to document the psychological processes that give rise to the judgment outcomes we observe.

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. Much of the variance in estimation and placement appear to be driven by task difficulty. We do not find consistent differences between people in their tendencies to overestimate or overplace their performances, as evidenced by their low alpha reliabilities.

We observe more systematic overprecision, as well as more consistent differences between people in its display. Furthermore, the tendency for greater precision in judgments is correlated with overestimation and overplacement. Greater precision in self-estimates appears to reduce overestimation. In addition, greater precision in other-estimates appears to reduce overplacement.

Our data do not reveal useful predictors of overconfidence among the individual-difference questionnaire measures of individual differences we used.

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 et al., 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.

False consensus and false uniqueness. Our results reveal the simultaneous co-occurrence of both false consensus (thinking that others are more like you than they are, L. Ross, Greene, & House, 1977) and false uniqueness (believing that others are more different from you than they are, Perloff & Brickman, 1982) on the same dependent measures (see also Moore & Kim, 2003; Moore & Small, in press-a). Participants' judgments reveal false uniqueness because participants reported believing that their scores were more exceptional than they actually were—.48 points higher than those of others on the easy quiz and 1.36 points lower on the hard quiz. At the same time, participants' judgments reveal false consensus because people believed that others would perform more like them than they actually did. Performance by participants and their RSPPs were uncorrelated within quiz (r = -.05). However, participants believed that others were like them—estimations of performance by self and other are strongly correlated within quiz (r = .48).

As Dawes (1989) pointed out, there is a rational basis for the so-called "false consensus" effect. If the only source of information participants had about quiz difficulty was their own performance, then they would do well to assume others would perform as they did (Krueger, 2003; Krueger et al., 2005). Indeed, people who assume others are like them will, on average, make more accurate estimates of others than will people who ignore the information they have about themselves and do not update their prior beliefs (Dawes & Mulford, 1996). And false uniqueness effects arise quite naturally from the fact that people possess better information about themselves than about others, as Fiedler (1996) has shown.

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; C. T. F. 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 (in press). 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 & Hoffrage, 1995). 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.

Absent/exempt. When people have exposed themselves to some risk but haven't experienced a negative outcome, they sometimes conclude that they are immune to that risk. For instance, young women who have unprotected sex and do not get pregnant may incorrectly conclude that they are infertile (Downs, Bruine de Bruin, Murray, & Fischhoff, 2004); or sexually active college students who have had unprotected sex but who haven't contracted AIDS may come to believe think they're less vulnerable to getting AIDS than are others (Weinstein, 1982). Basically, their personal experience with (the absence of) a rare negative event leads them to reduce their estimates of their own personal risk, and they wind up believing that they are less likely than others to experience it (see Hertwig, Barron, Weber, & Erev, 2004). People remain quite clueless about the low actual risk of contracting AIDS (roughly 1 in 500 chance of becoming infected from a single act of unprotected coitus with an infected partner, Hearst & Hulley, 1988), and so overestimate this small probability. This result is obviously 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.

Debiasing. Metcalfe's (1998) review discusses debiasing methods that work to reduce or eliminate overestimation where it does occur. These are all strikingly consistent with our differential information theory. Most of Metcalfe's recommendations involve giving people better information about themselves or others. Indeed, Metcalfe (1998) argues that what evidence exists for overestimation is more consistent with poor information about one's own performance (and the causes of it) than with self-deception.

Extensions to Other Phenomena

Fiedler's work (1991; 1996; 2000a) has persuasively demonstrated the broad applicability of explanations such as the one we present in this paper. Fiedler applied his BIAS model to help explain such diverse phenomena as outgroup homogeneity, ingroup-serving bias, confirmation bias and auto-verification, illusory correlations, the mere thinking effect on attitude polarization, group polarization, constructive memory illusions, unpacking and category-split effects, and numerous other applications. 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 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. Some have found that overestimation increases with the personal importance of the task domain (M. Ross, McFarland, & Fletcher, 1981; Sanbonmatsu et al., 1987; Sanitioso, Kunda, & Fong, 1990). Other research has found that people are more likely to believe that they are better than others with whom they expect to compete vs. cooperate (W. M. P. Klein & Kunda, 1992). However, it is not simply the case that overconfidence is driven by event desirability. A number of researchers have examined the effect of event desirability on optimism and found no effects (Helweg-Larsen & Shepperd, 2001; Weinstein, 1980). As Krizan and Windschitl’s (2007) review of the literature shows, the evidence for wishful thinking, in which optimism is influenced by event desirability, is not strong.

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).

While research on motivation has generally focused on motivational causes of overestimation and overplacement, we ought to point out the fact 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.

What constitutes performance? We have implicitly assumed that it is easy to specify what constitutes performance. In our study, there is little ambiguity about how many items each participant answered correctly on each trivia quiz. However, a substantial proportion of the evidence showing overplacement employs tasks on which performance is largely subjective. Svenson (1981), for instance, did not tell his participants exactly what constituted driving skill—he asked them how their driving skills compared with those of other drivers and left it up to them to determine how to assess driving skill. Indeed, Dunning, Meyerowitz, and Holzberg (1989) showed that the more subjective the domain, the more latitude people have to claim that they are better than others. This may be because the motivation to claim undeserved distinctiveness is unleashed in domains where people know that nobody can expose their exaggerations. It could also be the consequence of varied and idiosyncratic definitions of what constitutes performance (Santos-Pinto & Sobel, 2005; Van den Steen, 2004).

We should note that when the performance criterion is vague, it becomes difficult to assess the accuracy of estimations of performance. Consequently, it is not easy to measure the relationship between overestimation and overplacement in vague domains. This is why these vague performance domains did not provide a useful context in which to test our theory.

Direct vs. indirect measures. The leading explanation for above- and below-average effects is differential weighting: When making comparative judgments, people overweight self-knowledge and underweight knowledge about others (J. R. Chambers & Windschitl, 2004; Giladi & Klar, 2002). 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 (J. R. 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.)

The self-selection problem. Even if, as our theory implies, people believe just as easily in their own inferiority as in their superiority, it does not follow that we should expect to see unbiased judgment in everyday judgments. In life, people are not randomly assigned to the universe of all possible tasks. Instead, people self-select based, in part, on where they believe they can distinguish themselves from others (Tesser, 1988). This fact has two important consequences. First, easy tasks, on which more people feel competent, will attract too many competitors (Moore, Oesch, & Zietsma, in press). It is, after all, industries such as restaurants, bars, and hobby shops that see the highest rates of entry, most intense competition, and, consequently, the highest rates of failure (Geroski, 1996; U.S. Small Business Administration, 2003). Second, we should expect overplacement to be the rule if we sample the beliefs of only those who have chosen to enter or compete.

Future Research

Our failure to replicate prior findings regarding correlations between measures of individual difference and overconfidence is notable. We cannot account for this discrepancy. We can only speculate that its cause may have to do with differences between our measures of overconfidence and those used in those other studies. However, this is clearly a matter for future research.

We think that the greater robustness of overprecision in our data is interesting, but we are reluctant to conclude that overprecision is a universal tendency. Research clearly shows that overprecision is sensitive to the situational context (Block & Harper, 1991) and the elicitation method (Budescu & Du, in press; Juslin & Persson, 2002; Juslin et al., 1999). Furthermore, important questions remain about the causes for overprecision. Some research has suggested that anchoring and insufficient adjustment might be at work in causing overprecision (Block & Harper, 1991; Juslin et al., 1999). If people anchor on an initial answer and adjust insufficiently from it when building a confidence interval around it, that would produce apparent overprecision. Other researchers have suggested that confirmation bias or biased information retrieval may be at work (Hoch, 1985; Klayman et al., 1999; Koriat, Lichtenstein, & Fischhoff, 1980). Some research suggests the same memorial processes that give rise to confirmation bias may lead to anchoring (Mussweiler & Strack, 1999; Strack & Mussweiler, 1997), and maybe also to overprecision. Some of the most promising evidence highlights the role of faulty statistical intuitions surrounding the construction of confidence intervals (Juslin et al., in press).

Despite the many studies that that have shown overprecision, it has been examined primarily in just two ways: (1) using interval estimates of numerical answers to trivia questions and (2) using the item-confidence paradigm, which confounds it with overestimation.

Overprecision deserves to be studied with a greater variety of methods and in a greater variety of contexts than it has been in the past.

Conclusions

We began this paper by discussing findings on overconfidence that seem perplexingly 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 reduce both overestimation and 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.

Appendix A

Correlations across the 82 individuals between measures of overconfidence and measures of individual difference. Alpha reliability measures for the three different types of overconfidence also appear.

	Overestimation	Overplacement	Overprecision
Overestimation	$\alpha = .21$		
Overplacement	0.06	$\alpha = .29$	
Overprecision	0.24	-0.03	$\alpha = .95$
Extraversion	-0.15	-0.02	-0.08
Agreeableness	0.13	0.01	0.12
Conscientiousness	-0.02	-0.09	0.11
Neuroticism	-0.08	-0.15	0.04
Openness	-0.11	-0.26	-0.08
Social comparison orientation	0.03	0.14	0.05
Narcissisism	-0.01	0.09	-0.19
Self-esteem	-0.13	0.03	-0.10
Overclaiming	0.21	0.07	-0.12
Generalized sense of power	-0.16	0.01	-0.12
Perspective-taking	0.12	-0.11	-0.07
Empathy	0.13	0.00	-0.13
Cognitive reflection	-0.19	0.08	-0.19
Social conservatism	-0.10	-0.01	0.17
Economic conservatism	-0.10	-0.01	0.20
Male	0.18	0.13	0.02
Age	0.18	-0.08	0.10

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., John, O. P., & Keltner, D. (2005). *The subjective sense of power: Structure and antecedents*.

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*, 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.

Blanton, H., Axsom, D., McClive, K. P., & Price, S. (2001). Pessimistic bias in comparative evaluations: A case of perceived vulnerability to the effects of negative life events. *Personality and Social Psychology Bulletin, 27*(12), 1627-1636.

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*.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67*(3), 366-381.

Burson, K. A., & Klayman, J. (2005). *Judgments of performance: The relative, the absolute, and the in-between*. Ann Arbor: Unpublished manuscript.

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.

Christensen-Szalanski, J. J., & Bushyhead, J. B. (1981). Physicians' use of probablistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 928-935.

Clayson, D. E. (2005). Performance Overconfidence: Metacognitive Effects or Misplaced Student Expectations? *Journal of Marketing Education, 27*(2), 122-129.

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

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*(1), 1-73.

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

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113-126.

Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology, 25*(1), 1-17.

Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes, 65*(3), 201-211.

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

Dougherty, M. R. P. (2001). Integration of the ecological and error models of overconfidence using a multiple-trace memory model. *Journal of Experimental Psychology: General, 130*(4), 579-599.

Downs, J., Bruine de Bruin, W., Murray, P. J., & Fischhoff, B. (2004). When "it only takes once" fails: perceived infertility predicts condom use and STI acquisition. *Journal of Pediatric and Adolescent Gynecology, 17*(3), 224.

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

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57*(6), 1082-1090.

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

Fiedler, K. (2000b). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology, 4*(1), 25-58.

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.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42.

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.

Geroski, P. A. (1996). What do we know about entry? *International Journal of Industrial Organization, 13*(4), 421-441.

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology, 76*(1), 129-142.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental modules. In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 127-171). London: Routledge.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*(4), 684-704.

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.

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., & Buehler, R. (1999). Frequency, probability, and prediction: Easy solutions to cognitive illusions? *Cognitive Psychology, 38*(1), 48-78.

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*. Unpublished manuscript.

Hearst, N., & Hulley, S. B. (1988). Preventing the heterosexual spread of AIDS: Are we giving our patients the best advice? *Journal of the American Medical Association, 259*, 2428-2432.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review, 5*(1), 74-95.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*(8), 534-539.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 719-731.

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.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*, 1-55.

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

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.

John, O. P., & Strivastava, S. (2000). The big-five trait taxonomy: History, measurement, and theoretical perspectives. In L. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research*. New York: Guilford.

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.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129*(3), 339-375.

Juslin, P., Olsson, H., & Björkman, M. (1997). Brunswikian and Thurstonian origins of bias in probability assessment: On the interpretation of stochastic components of judgment. *Journal of Behavioral Decision Making, 10*(3), 189-209.

Juslin, P., & Persson, M. (2002). PROBabilities from EXemplars (PROBEX): A "lazy" algorithm for probabilistic inference from generic knowledge. *Cognitive Science, 26*(5), 563-607.

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.

Klein, W. M. P., & Kunda, Z. (1992). Motivated person perception: Constructing justifications for desired beliefs. *Journal of Experimental Social Psychology, 28*(2), 145-168.

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

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107-118.

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). The influence of outcome desirability on optimism. *Psychological Bulletin, 133*(1), 95-121.

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. (2003). Return of the ego--Self-referent information as a filter for social prediction: Comment on Karniol (2003). *Psychological Review, 110*(3), 585-590.

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.

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.

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist, 41*(9), 954-969.

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.

Metcalfe, J. (1998). Cognitive optimism: self-deception or memory-based processing heuristics? *Personality and Social Psychology Review, 2*(2), 100-110.

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. (in press). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior & Human Decision Processes*.

Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of Personality and Social Psychology, 85*(6), 1121-1135.

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., Oesch, J. M., & Zietsma, C. (in press). What competition? Myopic self-focus in market entry decisions. *Organization Science*.

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

Moore, D. A., & Small, D. A. (in press-b). 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.

Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35*(2), 136-164.

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. (in press). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*.

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.

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*(4), 890-904.

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.

Perloff, L. S., & Brickman, P. (1982). False consensus and false uniqueness: Biases in perceptions of similarity. *Academic Psychology Bulletin, 4*(3), 475-494.

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.

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Investory and and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54*(5), 890-902.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*(3), 279-301.

Ross, M., McFarland, C., & Fletcher, G. J. (1981). The effect of attitude on the recall of personal histories. *Journal of Personality and Social Psychology, 40*(4), 627-634.

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.

Sanitioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology, 59*(2), 229-241.

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. (2000). Rejoinder: the perils of Viscusi's analyses of smoking risk perceptions. *Journal of Behavioral Decision Making, 13*(2), 273-276.

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.

Sniezek, J. A., & Buckley, T. (1991). Confidence depends on level of aggregation. *Journal of Behavioral Decision Making, 4*, 263-272.

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.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*(2), 161-188.

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.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73*(3), 437-446.

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.

Tesser, A. (1988). Toward a self evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 181-227). New York: Guilford.

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

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273-286.

U.S. Small Business Administration. (2003). *Longitudinal Establishment and Enterprise Microdata*. Washington, DC: Office of Advocacy (202-205-6530).

Van den Steen, E. (2004). Rational overoptimism (and other biases). *American Economic Review, 94*(4), 1141-1151.

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.

Viscusi, W. K. (2000). Comment: The perils of qualitative smoking risk measures. *Journal of Behavioral Decision Making, 13*, 267-271.

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

Weinstein, N. D. (1982). Unrealistic optimism about susceptibility to health problems. *Journal of Behavioral Medicine, 5*(4), 441-460.

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

Wells, G. L. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology, 62*(5), 739-752.

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.

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.

Zenger, T. R. (1992). Why do employers only reward extreme performance? Examining the relationships among performance, pay, and turnover. *Administrative Science Quarterly*, *37*, 198-219.