

BAYESIAN OVERCONFIDENCE

DON A. MOORE
Carnegie Mellon University
don.moore@alumni.carelon.edu

PAUL J. HEALY
The Ohio State University
healy.52@osu.edu

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 those of others, and (3) overprecision in beliefs about the accuracy of one's knowledge.

INTRODUCTION

Overconfidence has long been recognized as a significant bias with meaningful organizational and economic consequences. We deal with three types of overconfidence: (1) overestimation (overestimating one's actual performance), (2) overplacement (the belief that one is better than others), and (3) overprecision (having beliefs which are overly precise).

Recent evidence suggests that these different measures may be related in an unexpected way: contexts that produce overplacement also tend to produce underestimation, and vice-versa (Moore & Kim, 2003; Moore & Small, in press). On easy tasks, people tend to believe that they are better than others (Alicke & Govorun, 2005) but underestimate how well they have done (Burson, Larrick, & Klayman, 2006). On difficult tasks, people tend to believe that they are worse than others (Kruger, 1999) but overestimate how well they have done (Erev, Wallsten, & Budescu, 1994). We ask whether these results stem from true biases in judgment, or whether they are consistent with Bayes's Law for an appropriately-defined inference problem. We present an experiment that replicates these effects and shows that the results are very much in line with Bayes's Law when agents use their own experience as an imperfect signal of the experience of others.

Stated simply, our theory is this: After experiencing a task, people 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.

In what follows, we elaborate and formalize this theory.

Bayesian Overconfidence

In this section we show how Bayesian inference can generate the patterns of overconfidence and underconfidence observed in previous research. We assume that each agent i performs an identical task in isolation and receives a numerical score, denoted x_i , which quantifies their performance. Each agent believes that each score x_i is a realization of a random variable X_i , and X_i is determined by (1) $X_i = S + L_i$, where S is the overall expected score across agents (or, the *simplicity* of the task) and L_i is a mean-zero ‘*luck and ability*’ variable. Assuming agent i has well-defined prior beliefs about the distributions of S , A_i , and L_i , she can update those beliefs upon observing her realized score x_i . If agent i also has well-defined priors on L_j for some other agent $j \neq i$, her beliefs about X_j will also change as she observes x_i and updates her belief about S . In this way agent i may overplace or underplace her score relative to others’ after she observes her own score.

More formally, suppose that before performing the task, i believes S has density π , A_j has density f_j for each agent j , and each L_j is independently and identically distributed with density g . Let $E[S] = \mu$. (We denote the expected value of a random variable with respect to i ’s beliefs by $E[\cdot]$.)

Overplacement. Under the above assumptions, i ’s prior expectations of her own score (with no experience) and the score of another agent are $E[X_i] = E[X_j] = \mu$. After performing the task and observing x_i , she updates her beliefs about S and L_i . Since she does not (yet) observe x_j , her beliefs about L_j remain unchanged, so that $E[X_j | x_i] = E[S | x_i]$. We say that i exhibits *overplacement* if $E[X_j | x_i] < x_i$ and *underplacement* if $E[X_j | x_i] > x_i$.

Suppose an agent i who has never encountered the task before receives a score higher than expected ($x_i > \mu$). She might infer that her high score was due to good luck ($l_i > 0$) or a simpler-than-expected task ($s > \mu$). If she attributes her high score entirely to the task’s simplicity (*i.e.*, she believes $E[S | x_i] = x_i$), then she will exhibit no overplacement because task simplicity affects all agents equally. If instead she attributes her high score at least partially to her own luck (*i.e.*, she believes $E[S | x_i] < x_i$), then she will exhibit overplacement since $E[X_j | x_i] = E[S | x_i] < x_i$. Similarly, if $x_i < \mu$ and she attributes her low score at least partially to luck, then she will exhibit underplacement. The overplacement and underplacement phenomenon therefore hinges on the specification of agents’ prior beliefs and the resulting behavior of $E[S | x_i]$.

This model serves as the basis for two hypotheses:

Hypothesis 1. When a task is easier than expected, people will believe that they are better than others.

Hypothesis 2. When a task is more difficult than expected, people will believe that they are worse than others.

Overestimation. In many situations, the act of performing a task does not perfectly reveal one’s performance. A new motorist completing a driving test, a runner crossing the finish line, and a high school student turning in her standardized test booklet all have a rough idea of how well they performed, but they do not know their score with certainty. The uncertainty is eliminated only after the driving instructor announces the result, the runner glances at the scoreboard, or the anxious student receives her score in the mail.

Formally, we assume that after performing a task, agent i does not know her true score x_i , but observes a draw of the random variable $Y_i = x_i + E_i$, where E_i is a random variable with a symmetric, mean-zero distribution that may depend on x_i . Since we assume $X_i = S + L_i$, an agent who observes a draw y_i from Y_i makes inferences about S , L_i , and E_i . For example, a high value of y_i may lead her to conclude that S is relatively high, but that L_i and E_i were positive as well. In this case, she will expect that she did better than average (because L_i is positive), but not as good

as her signal (because E_i is positive.) If her signal is in fact accurate, then she has underestimated her actual performance. Formally, we say that i exhibits *overestimation* if $E[X_i|y_i] > x_i$ and *underestimation* if $E[X_i|y_i] < x_i$.

This model serves as the basis for two further hypotheses:

Hypothesis 3. When a task is easier than expected, people will underestimate their own performances.

Hypothesis 4. When a task is more difficult than expected, people will overestimate their own performances.

Overprecision. If a person's beliefs about her own score has lower variance (higher precision) than her actual distribution of scores, we say that she exhibits *overprecision*. Since our description of agents' inferences operates only on subjective beliefs without assuming those beliefs are empirically accurate, the presence of overprecision will not qualitatively affect the above results on overplacement and overestimation; however, overprecision may affect the magnitudes of overplacement and overestimation.

EXPERIMENTAL DESIGN

Eighty-two student participants were recruited from the undergraduate population of Carnegie Mellon University and participated on computer terminals in the Center for Behavioral Decision Research laboratory. In each of 18 rounds, each participant completed a 10-item trivia quiz. Each participant earned $\$25r$ in each round, where r is her percentile rank on the quiz relative to all other participants who had already taken the same quiz, plus payments earned from a series of predictions made about their own scores and the scores of other participants. 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 make predictions of the probability that they would obtain each of the eleven possible scores (0 through 10.) Subjects were paid for their predictions using the quadratic scoring rule, which maximized each participant's payoff when he or she made the best estimate of the actual distribution of possible scores (Selten, 1998). In addition to announcing her prior belief on her own score, each subject was asked to announce a prior belief about the score of a randomly-selected previous participant (RSPP) who had previously taken the same quiz. This prediction was also incentivized by the quadratic scoring rule. After submitting prior distributions for 'self' and 'other', subjects took the 10-question quiz. Before learning their quiz score, subjects were again asked to predict their own score and the score (on the same quiz) of the same RSPP. Next, each subject was shown the correct answers and graded their own quiz. Finally, subjects were asked once again to predict the score of the RSPP. At the end of the round each subject was shown the RSPP's score, her percentile rank r , and her payoff from the five scoring rules. The round was then repeated with a new 10-item quiz and a new RSPP for each subject. After all 18 rounds were complete, subjects were paid based on their performance in five randomly chosen rounds. The 18 quizzes span six topics, each at three difficulty levels. The quizzes were randomly assigned to six three-round blocks such that each block had one quiz of each difficulty level.

RESULTS

For each participant in each round we observe five probability distributions: The person's prior and interim beliefs about her own score and her prior, interim, and posterior beliefs about

the score of the RSPP. The averages of the expected values of these distributions (across all players and periods) for each quiz difficulty level are given in Table 1. Note that actual scores appear in the table under the posterior phase.

Table 1. Averages (and standard errors) of expected values of reported belief distributions.

Difficulty	Prior		Interim		Posterior	
	Own	RSPP	Own	RSPP	Actual	RSPP
Easy	5.16 (1.6)	5.15 (1.2)	8.64 (2.2)	8.26 (1.6)	8.86 (2.2)	8.50 (1.5)
Medium	5.30 (1.6)	5.34 (1.3)	5.93 (3.1)	6.15 (2.1)	5.92 (3.2)	6.21 (2.1)
Hard	5.63 (1.7)	5.58 (1.3)	1.50 (1.7)	2.95 (1.9)	0.71 (1.4)	2.35 (1.9)

We now demonstrate two main results using the data: First, participants show overplacement on easy quizzes and underplacement on hard quizzes. Second, participants underestimate their scores on easy quizzes and overestimate them on hard quizzes. These results are demonstrated by regressions whose estimates appear in Table 2. In each regression, an appropriate dependant variable is regressed against a full set of dummy variables indicating easy, medium, and difficult quizzes. Each regression was also run with dummy variables for block effects and all interactions between blocks and difficulties.

Table 3. Dummy variable regressions demonstrating the four main results. Superscripts indicate interim expectations (1) and posterior expectations (2). “Score” refers to the participant’s own score. Bold-faced entries are significant at the 5% level.

Result	1		2
Dependant Variable	$E^1(\text{Self}) - E^1(\text{Other})$	Score - $E^2(\text{Other})$	$E^1(\text{Self}) - \text{Score}$
Easy	0.385 (4.02)	0.369 (3.61)	-0.219 (-4.11)
Medium	-0.221 (-2.30)	-0.284 (-2.78)	0.006 (0.11)
Difficult	-1.448 (-15.12)	-1.701 (-16.67)	0.851 (15.92)

Result 1: *Participants overplace their scores after easy quizzes and underplace their scores after difficult quizzes.* Columns 2, 3, and 4 of Table 2 test our hypotheses. We examine two measures of overplacement: interim-phase overplacement and posterior-phase overplacement. The first represents overplacement while participants are uncertain about their own scores and the second represents overconfidence after this uncertainty is resolved. The regression in column 2 indicates that participants exhibit significant overplacement in the interim phase after an easy quiz and significant underplacement after a difficult or medium quiz. Specifically, participants expect to out-perform the RSPP by an average of 0.39 points after an easy quiz, but expect to be out-performed by an average of 1.45 points after difficult quizzes. The result is similar in the posterior phase; participants overplace their scores by an average of 0.37 points after easy quizzes and underplace by 1.70 points after difficult quizzes.

Result 2: *Participants underestimate their scores after easy quizzes and overestimate their scores after difficult quizzes.* Our measure of overestimation is the difference between a participants’ expected score in the interim phase (after taking the quiz) and their actual score. The final regression from Table 3 confirms the predictions of the theory. Participants underestimate their scores by 0.22 points after easy quizzes and overestimate them by 0.85 points after difficult quizzes. Overestimation on medium quizzes is essentially absent on medium quizzes.

Overprecision. Recall that overprecision occurs when agents' belief distributions have lower variance than the distribution of actual outcomes. When comparing the actual variance of scores on that particular quiz with the variance of participants' post-quiz reports about others' performance, we see that the former is on average larger by 2.3. After participants learn their own scores, this difference increases insignificantly to 2.5. Both are significantly greater than zero, indicating the presence of overprecision in post-quiz beliefs.

DISCUSSION

There have been a number of recent models that have attempted to explain how rational Bayesian agents could display overconfidence (Benabou & Tirole, 2002; Bodner & Prelec, 2003; Van den Steen, 2004). None of these models can parsimoniously account for the evidence from the present experiment because they do not predict the systematic underconfidence observed in the experimental results.

These previous models were built in an attempt to account for the systematic findings of overconfidence in the empirical literature. The question then arises how it is that the present experiment has found what appears to be a unique exception to such systematic empirical regularities. The answer to this question has two parts. The first part of the answer is that the majority of studies finding overestimation have done so in a way that confounds overestimation with overprecision (Alba & Hutchinson, 2000; Fischhoff, Slovic, & Lichtenstein, 1977), making it impossible to determine the degree to which each is responsible for the result.

The second part of the answer is that the strongest previous findings of relative and overestimation have not occurred in the same studies. Those studies in which people overestimate their absolute performance the most have tended to focus on contexts in which performance is low and success is rare (Juslin, Winman, & Olsson, 2000; Malmendier & Tate, 2005; Weinstein, 1980). Those studies in which people overestimate their relative performance the most have tended to focus on context in which performance is high and success is likely (Kruger, 1999; Messick, Bloom, Boldizar, & Samuelson, 1985; Svenson, 1981).

This paper proposes an integration of inconsistent empirical results in studies of overconfidence. The evidence suggests that the three ways in which overconfidence has previously been viewed are distinct from one another and should not be regarded as the same phenomenon. The tendency to overestimate one's actual performance is greatest on difficult tasks. The tendency to overplace one's performance relative to others is greatest on simple tasks. One implication of this argument is that the third variety of overconfidence, the tendency to overestimate the quality of one's own private signal (or the precision of one's beliefs) is separate from the first two and may reduce their magnitudes.

There are other testable implications of the present theory. One is that in simple competitions will elicit greater interest and more entrants than will difficult competitions. Evidence on rates of entrepreneurial entry suggest a remarkably high correlation between entry and exit across industries over time (Moore & Cain, in press). These patterns are not well explained by inter-industry differences in rates of profitability or barriers to entry (Geroski, 1996). It is possible that "simple" industries—for which many people think they would know how to run a business—such as restaurants, bars, clothing retail, and liquor stores—see higher rates of entry, more intense competition, and higher rates of failure.

An important implication of the present study is a methodological one. It is a mistake to assume that all three types of overconfidence represent the same underlying psychological bias. They are distinct from each other. Empirically, overestimates of absolute performance and

overplacement relative to others tend to be negatively correlated across tasks. And the tendency to overestimate the quality of one's private signals should decrease both of the other types of biases. Those who study overconfidence and its implications must be explicit about which form they are studying.

REFERENCES

- 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., & 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.
- Benabou, R., & Tirole, J. 2002. Self-confidence and personal motivation. *Quarterly Journal of Economics*, 117: 871-915.
- Bodner, R., & Prelec, D. (2003). Self-signaling and diagnostic utility in everyday decision making. In I. Brocas & J. D. Carillo (Eds.), *The psychology of economic decisions, volume 1: Rationality and well-being* (pp. 105-123). Oxford: Oxford University Press.
- 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: 60-77.
- Erev, I., Wallsten, T. S., & Budescu, D. V. 1994. Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review*, 101: 519-527.
- Fischhoff, B., Slovic, P., & Lichtenstein, S. 1977. Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3: 552-564.
- Geroski, P. A. 1996. What do we know about entry? *International Journal of Industrial Organization*, 13: 421-441.
- 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: 384-396.
- Kruger, J. 1999. Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77: 221-232.
- Malmendier, U., & Tate, G. 2005. CEO overconfidence and corporate investment. *Journal of Finance*: 6.
- Messick, D. M., Bloom, S., Boldizar, J. P., & Samuelson, C. D. 1985. Why we are fairer than others. *Journal of Experimental Social Psychology*, 21: 480-500.
- 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: 1121-1135.
- Moore, D. A., & Small, D. A. in press. Error and bias in comparative social judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology*.
- Selten, R. 1998. Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, 1: 43-61.
- Svenson, O. 1981. Are we less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47: 143-151.
- Van den Steen, E. 2004. Rational overoptimism (and other biases). *American Economic Review*, 94: 1141-1151.
- Weinstein, N. D. 1980. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39: 806-820.