

Running head: OVERCONFIDENCE AND UNDERCONFIDENCE

Overconfidence and Underconfidence:

When and Why People Underestimate (and Overestimate) the Competition

Don A. Moore  
Carnegie Mellon University  
5000 Forbes Avenue  
Pittsburgh, PA 15213  
412-268-5968  
Fax: 412-268-7345  
dmoore@cmu.edu

Daylian M. Cain  
Harvard University  
1805 Cambridge Street  
Cambridge, MA 02138  
617-678-2246  
Fax: 617-495-7730  
cain@fas.harvard.edu

NOTE: This article has appeared in print. This is not the copy of record. See: Moore, D. A. & Cain, D. M. (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes*, 103, 197-213.

\* The authors appreciate the insightful comments, on earlier versions on this manuscript, by Linda Babcock, J. Nicolas Barbic, Max Bazerman, Jason Dana, Paul Geroski, PJ Healy, Chip Heath, Erik Hoelzl, George Loewenstein, Daniel Lovallo, Rob Lowe, John Oesch, John Patty, Vahe Poladian, Jesper Sorensen, Lise Vesterlund, and Roberto Weber. Thanks to Sapna Shah and Sam Swift for help with data collection. The authors also appreciate the support of National Science Foundation Grant SES-0451736, a Berkman Faculty Development Grant at Carnegie Mellon, and the assistance of John Duffy in the use of the Pittsburgh Experimental Economics Laboratory at the University of Pittsburgh for collecting the experimental data. Address correspondence to: don.moore@alumni.carleton.edu.

Abstract

It is commonly held that people believe themselves to be better than others, especially for outcomes under their control. However, such overconfidence is not universal. This paper presents evidence showing that people believe that they are *below* average on skill-based tasks that are difficult. A simple Bayesian explanation can account for these effects and for their robustness: On skill-based tasks, people generally have better information about themselves than about others, so their beliefs about others' performances tend to be more regressive (thus less extreme) than their beliefs about their own performances. This explanation is tested in two experiments that examine these effects' robustness to experience, feedback, and market forces. The discussion explores the implications for strategic planning in general and entrepreneurial entry in particular.

Key words: Entrepreneurial entry, overconfidence, controllability, skill, competence, entrepreneurship, better-than-average, reference group neglect, egocentrism, differential regression, comparative judgment

## Overconfidence and Underconfidence:

### When and Why we Overestimate (and Underestimate) the Competition

One of the most popular social psychology textbooks states, "For nearly any subjective and socially desirable dimension...most people see themselves as better than average" (Myers, 1998, p. 440). For example, people report themselves to be above average in driving ability, their ability to get along with others, and their chances of obtaining jobs that they like (College Board, 1976-1977; Svenson, 1981; Weinstein, 1980). Some have argued that the most important business decisions, including the decision to found a new firm, enter an existing market, or introduce a new product are routinely biased by such overconfidence (Cooper, Woo, & Dunkelberg, 1988; Dunning, Heath, & Suls, 2004; Hayward & Hambrick, 1997; Malmendier & Tate, 2004; Odean, 1998; Zajac & Bazerman, 1991).

Recent evidence, however, has cast doubt on the generality of overconfidence. There are a number of different domains in which people are systematically underconfident. For example, people believe that they are *below* average in unicycle riding, computer programming, and their chances of living past 100 (Kruger, 1999; Kruger & Burrus, 2004). It turns out that people tend to predict that they will be better than others on easy tasks where absolute performance is high, but worse than others on difficult tasks where absolute performance is low (Hoelzl & Rustichini, 2005; Moore & Kim, 2003; Windschitl, Kruger, & Simms, 2003). A number of researchers have explained this effect as egocentrism: People focus on their own performances and neglect consideration of others' (Camerer & Lovallo, 1999; Kruger, 1999). In this paper, we present a

new explanation for these better-than-average (BTA) and worse-than-average (WTA) effects.<sup>1</sup>

Our explanation holds that BTA and WTA effects are a natural consequence of regressive estimates of others, which result from the fact that people have better information about themselves than they do about others. We test this explanation using two experiments that examine the robustness of BTA and WTA effects to experience, feedback, and market forces. The results are consistent with our hypotheses, and have some provocative implications.

For the sake of exposition, let us introduce our theory by considering beliefs about performance on a one-question test where the answer is either right or wrong. Before having seen the problem, and without any information regarding its ease or difficulty, how likely are you to solve it? One assumption might be that performance will be uniformly distributed across possible outcomes (Fischhoff & Bruine De Bruin, 1999; Fox & Rottenstreich, 2003), leaving a 50% chance that you will solve the problem. Such an "ignorance prior" might make sense in the absence of better information. Whatever it is, this prior is simply your baseline expectation for your performance.

After taking the test, let us say that you know whether you solved the problem. What are you to believe about others' performances? If your own performance is useless for predicting others' (e.g., if you think that your good performance was based entirely on luck), your estimation of others' performances ought to remain unchanged from your prior beliefs. Therefore, doing well should leave you thinking that you did better than others; and doing poorly should leave you thinking that you did worse than others. Even if your beliefs about your own performance are helpful for predicting others', so long as there remains uncertainty about others' performances, your predictions of them should depend on—and thus regress towards—the

---

<sup>1</sup> We use the terms better- and worse-than-average to be consistent with prior work. We acknowledge that with skewed distributions, it is indeed possible for the majority of people to be above (or below) average. This concern, while valid, does not represent a problem for the results of the experiments we present.

ignorance prior. The upshot is that, when your absolute performance is better (or worse) than your prior expectations, sensible Bayesian inference will lead you to make predictions of others' performances that are between these priors and your current beliefs about your performance.

It is simple to extend this logic to a multi-item test: If one begins with the assumption that one is just as likely as others to get any given item correct, after having taken the test, one should estimate that others tend to score somewhere between one's own score and one's prior expectation. For example, suppose you initially expected everyone to score about 70%, but you think you scored about 90%. Depending on how indicative you feel your score is of others' scores, you might predict others to score, say, 80%. If you scored 50%, you might predict others to score, say, 60%. Notice that this perspective does not imply a belief in differences of overall ability between you and others—across both tests you would predict the same average score for everyone, namely 70%. But, on each test, you would be right to expect differences between you and others, given better information about your own score on that test. For a more formal development of this differential regression theory, see Appendix A.

Naturally, if the task includes no skill component whatsoever and performance is yet to be determined entirely by chance factors or factors outside one's control, then there would be little reason for people, on average, to predict that they would be above or below average. Consistent with this reasoning, a number of researchers studying BTA effects have found that they tend to be stronger on controllable tasks than on uncontrollable tasks (for a review, see Harris, 1996). For instance, Camerer and Lovallo (1999) found that potential market entrants were excessively confident about winning when competition was based on their skill but not when winners were selected randomly. The authors used this evidence to claim that high rates of entrepreneurial entry might be attributable to entrepreneurial overconfidence. However, because prior studies have employed *easy* tasks, the conclusion that people believe they are better than

average on all skill-based tasks is unwarranted. Instead, our theory would predict WTA effects when the task is more difficult than expected. We test this prediction in our first experiment. The first experiment also tests our theory that BTA and WTA effects are attributable to the differential regressiveness in estimates of self versus others. Experiment 2 addresses some shortcomings of Experiment 1 and provides further support for our theory that better information about self than others produces differential regressiveness.

### EXPERIMENT 1: THE MARKET ENTRY GAME

Our design builds on that of Camerer and Lovallo (1999). They devised an  $N$ -player coordination game in which, in each round,  $N$  players decide simultaneously and without communication whether to enter a market or not. Each market had a pre-announced capacity,  $c$ , which determined how many entrants earned money: Entrants ranked below  $c$  lose money, entrants ranked  $c$  or above earned money, while non-entrants neither earned nor lost any money. Each entrant's payoffs depended on his or her rank within the market, such that more money was earned by better performance relative to other entrants.

Camerer and Lovallo's key contribution over prior market-entry experiments was manipulating whether rankings were determined by either (a) a chance device, or (b) the entrant's skill (in answering trivia questions). This manipulation was implemented within-subject, so the same participants saw several rounds in which entrants were ranked based on a skill-based task and several rounds in which entrants were ranked randomly. They found that skill-dependent payoffs encouraged overconfidence and excess entry. Furthermore, excess entry was highest in sessions for which it was common knowledge that all participants were trivia enthusiasts, suggesting that participants were neglecting consideration of the reference group (similar enthusiasts) with which they would be competing.

However, our theory predicts underconfidence and insufficient entry as well. To test this, the new feature of our design is that skill-dependant payoffs are based on either an *easy* or a *difficult* trivia game. Contrary to the notion that overconfidence tends to be pervasive on all skill-based competitions, we predict that participants will only believe they are better than others on simple tasks, and thus, we expect excess entry only there.

We also test Camerer and Lovallo's explanation: that people focus on themselves and simply *neglect* consideration of others (rather than miscalculating others' performance) when making comparative judgments. Camerer and Lovallo called this "reference group neglect" and others have simply called it egocentrism (Chambers & Windschitl, 2004; Kruger, 1999). For example, as examinations become more difficult, students become more pessimistic about their final grades, even when it is common knowledge that the test will be graded on a forced curve (Windschitl et al., 2003). While our differential regression explanation would predict the same effect, the reference group neglect explanation holds that such false pessimism arises because students neglect to consider the fact that other students are also likely to find the test difficult. In other words, students trying to estimate their curved grades put too much weight on their own absolute performances. The differential regression explanation, on the other hand, hypothesizes that, regardless of the weighting attached to it, estimates of others will be more regressive than estimates of self. In summary, reference group neglect is about errors in the weight one puts on estimates of others' performance, while differential regression is about errors in the estimate that is weighed. We will measure the differential weighting hypothesized by reference group neglect, as well as other plausible alternative explanations, and show that the differential regression hypothesized by our theory can better account for our results.

The design of Experiment 1 includes several features that should help people avoid the mistake of ignoring or neglecting the competition: First, competitors are physically present,

salient, and individuated. Second, participants engage in a series of competitions over several rounds with full feedback each round, giving them the opportunity to learn.

### Method

In each round of our experiment, all seven participants in each experimental session were ranked relative to each other, according to a pre-announced method. Before the rankings were made public, we asked participants whether or not they wanted to enter into a competition in which only the three top-ranked entrants would make money. After they decided whether to enter, participants answered a number of questions regarding their own performances and the performances of others. Finally, participants received full feedback regarding absolute performances (of self and others), how many participants chose to enter each round, and the relative rankings of all (anonymously identified) entrants. The entire process was repeated over 12 rounds, with the three ranking methodologies (scores on a simple trivia quiz, scores on a difficult trivia quiz, or randomly generated scores) manipulated within session between rounds.

There were 13 experimental sessions, each with 7 people for a total of 91 individual participants. Participants were students at Carnegie Mellon University. Each participant was endowed with \$10. In each of the 12 rounds, participants decided whether to enter the market or whether to stay out and risk nothing. Entering the market meant either a loss or a gain, based on the entrant's rankings within that market. These payoffs are shown in Table 1.

The system by which players were ranked was announced publicly at the beginning of each round. In four of the twelve rounds, rankings were determined randomly. After they decided whether to enter, participants were assigned a randomly generated score from the set of real numbers between 0 and 5, inclusive. In the remaining eight (skill-based) rounds, rankings were based on trivia quizzes taken at the beginning of the round. Quizzes had five questions and a sixth tiebreaker question. Four of these eight rounds' quizzes were simple (with a mean score



of 4.58 out of 5) and four were difficult (with a mean score of .41 out of 5). The tiebreaker questions were scored based on the answer's distance from the correct numerical answer. The presence of this tiebreaker question virtually eliminated the chance of a tied score (there were none). The four simple and four difficult quizzes appear in Appendix B.

**Table 1.** Payoffs as a function of number of entrants and market rank. The table shows how much money was paid out in total (column 4) and per entrant (column 5).

<u>Rank</u>	<u>Payoff</u>	<u>Cumulative entrants</u>	<u>Cumulative payoff</u>	<u>Cumulative expected payoff per entrant (assuming ignorance about rankings)</u>
1 <sup>st</sup>	\$14	1	\$14	\$14
2 <sup>nd</sup>	\$10	2	\$24	\$12
3 <sup>rd</sup>	\$5	3	\$29	\$9.67
4 <sup>th</sup>	-\$10	4	\$19	\$3.50
5 <sup>th</sup>	-\$10	5	\$9	\$1.80
6 <sup>th</sup>	-\$10	6	-\$1	-\$0.17
7 <sup>th</sup>	-\$10	7	-\$11	-\$1.57

In order to rule out idiosyncratic effects of order, we varied the sequence in which participants encountered the three different ranking systems as follows. The different ranking systems (R= random, S=simple, and D=difficult) were arranged in three different orders which varied across experimental sessions: RSD, DRS, and SDR. Whatever sequence was arbitrarily chosen for the session was repeated four times, making twelve rounds in four three-round blocks. So, for example, if the first three rounds used the sequence RSD, all participants in that session faced the same quizzes at the same time, with ranking systems in the order: RSD-RSD-RSD-RSD. The order in which participants encountered the four different simple and difficult trivia quizzes was also counterbalanced between experimental sessions.

In the eight skill-rank rounds, after taking the given quiz but before seeing their scores, all participants simultaneously made their entry decisions (to enter or stay out). In the four random-rank rounds, there were no quizzes, and all participants merely made their entry decisions prior to learning their ranks. In all 12 rounds, after participants made their entry decisions they then answered the following questions:

1. *How many people total do you think will enter the market this round? Include yourself in this figure if you chose to enter.*
2. *What percentage of the other six entrepreneurs in this round do you think will score lower than you will (regardless of whether anyone enters)?*
3. *How many questions (out of 5) do you think you got correct this quiz? In random-rank rounds, this question was replaced with the question: What score (out of 5) do you think you will get this round?*
4. *How many questions (out of 5) do you think the average participant will get correct this round? In random-rank rounds, this question was replaced with the question: What score (out of 5) do you think the average participant will get this round?*
5. *If you chose to enter the market this round, what rank do you think you will get?*

At the end of each round, participants received full feedback regarding each of the seven players' individual scores, entry decisions, and rankings. In the eight skill-rank rounds, these scores were their trivia quiz performances; in the four random-rank rounds, these were their randomly generated scores. This information was posted using anonymous participant numbers on a blackboard in view of all participants. Each participant knew his or her own number, but did not know how the other numbers corresponded to those individuals present. Each of the 12 rounds'

results was left up for the entire experimental session. Participants' prior expectations regarding difficulty were not measured in this experiment, but they are in the second experiment.

At the end of the 12 rounds, three rounds were chosen at random to determine payoffs. The payoffs for these three rounds were averaged, and this amount was added to (or subtracted from) participants' \$10 endowment. Thus, the maximum possible payoff was \$24 for a participant who entered and was ranked first in each of the three payoff rounds (\$10 endowment plus an average of \$14 in total over the three selected payoff rounds). It was also possible for a participant to leave the experiment empty-handed if he entered and was ranked 4<sup>th</sup> or below on each of the three payoff rounds (\$10 endowment minus an average loss of \$10 in total). Across all participants, the mean final payoff was \$13.01 (with a range of \$4 to \$24).

*Equilibrium Predictions.* As Table 1 (column 5) shows, entry has a positive expected value so long as five or fewer players enter the market, assuming players have no information about their own relative rank. If players are risk-neutral, then it is rational (i.e., there is a set of pure-strategy Nash equilibria) for five players to enter each round. Lacking some coordinating device for deciding which of each session's seven total players enter and which stay out, there is a rational strategy (i.e., a mixed-strategy equilibrium) that is somewhat more complicated to compute, but the result is that all players enter with a probability of 84%. Naturally, since only the top three ranks actually win money, if all players know what their ranks will be, then only the top three players (3/7 or 43% of the potential entrants) will enter. Therefore, if all players were unbiased and imperfectly informed, we should expect between 43% and 84% of participants to enter each round.

Predicting the equilibrium outcome without the assumption of risk neutrality is more difficult. Even without information on their rankings, if everyone was sufficiently risk averse, no one would enter, and if everyone was sufficiently risk seeking, everyone would enter. So,

following Camerer and Lovallo (1999), we use the random-rank condition (when players cannot possibly have useful information on their rankings) to provide an empirical estimate of behavior given players' risk preferences. Deviations from entering 84% of the time in random-rank conditions suggest particular deviations from risk neutrality. And since all participants see all conditions, their entry decisions in the different conditions serve as within-subject controls for risk preferences. The difference in entry rates *between* the different conditions (random, simple, and difficult) is the dependent measure of primary interest.

*Hypotheses.* Consistent with the differential regression explanation, we predict that participants will believe themselves to be above average (and above median) on simple tests but below average (and below median) on difficult tests. As a result, we predict that participants will enter too frequently on simple-rank rounds and too rarely on difficult-rank rounds. We will take entry rates into random-rank rounds as indicators of participants' behavior given ignorance about their relative ranks and given their risk preferences. We predict that entry rates in random-rank rounds will lie between entry rates in simple- and difficult-rank rounds.

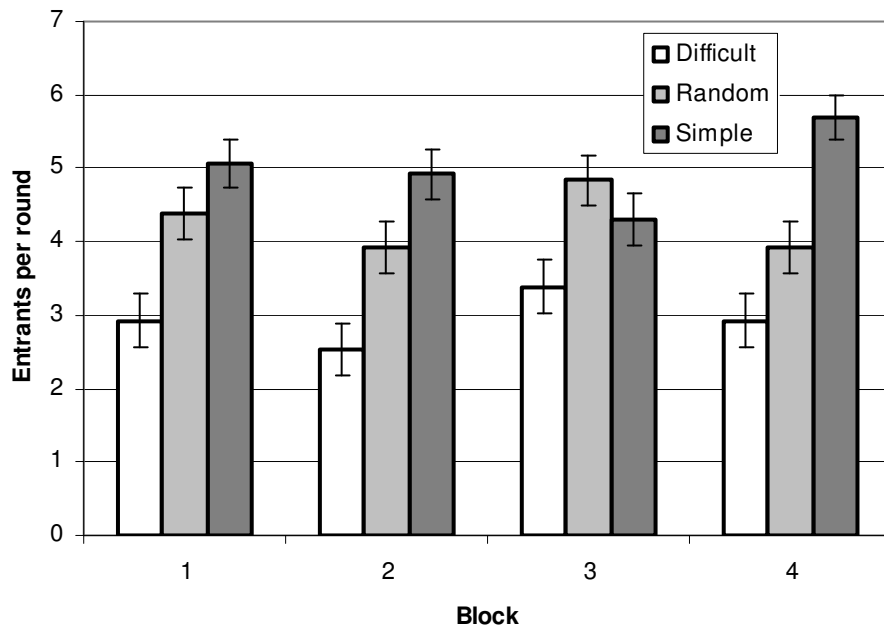
## Results

The average random-rank round saw 4.27 entrants (61% entry rate—suggesting slight risk aversion, on average). In contrast to this baseline, the average simple-rank round saw 5.0 entrants and the average difficult-rank round saw 2.94 entrants. In order to test for the statistical significance of these differences, we conducted a  $(4) \times (3)$  within-subjects ANOVA in which each of the 13 experimental sessions served as a single independent case. The four three-round blocks served as the first within-subjects factor and the three ranking systems served as the second within-subjects factor. The results reveal a significant effect of the experimental condition,  $F(2, 24) = 39.17, p < .001, \eta^2 = .77$ . Contrast tests confirm the significance of both

the difference between entry rates in difficult and the random-rank rounds ( $p < .001$ ) and the difference between the simple- and random-rank rounds ( $p = .018$ ).

The main effect of block is not significant,  $F(3, 36) = .82, p = .49, \eta^2 = .06$ . Although the interaction between block and ranking system (as shown in Figure 1) is marginally significant in the overall ANOVA,  $F(6, 72) = 2.08, p = .07, \eta^2 = .15$ , this does not appear to result from a consistent reduction in the effect of ranking system as participants gained experience: Entry rates in the difficult, random, and simple markets were 2.9, 4.4, and 5.1 respectively in the first block and were similarly 2.9, 4.2, and 5.7 respectively in the last block. Figure 1 shows these means.

**Figure 1.** Entry rates in the three different ranking systems across the four blocks. Error bars show standard errors.



*Explaining differences in rates of entry.* There are four possible explanations for the systematic effect of ranking systems on rates of entry: our differential regression explanation and three alternatives. The first alternative explanation is that participants believed that others would stay out of simple-rank rounds and so entry would have a higher expected value in simple

rounds. The data contradict this explanation: Participants predicted that there would be 5.2 entrants in the average simple-rank round (there were 5), 4.5 entrants in the average random-rank round (there were 4.27), and 3.2 entrants in the average difficult-rank round (there were 2.94). These predictions are consistent over time and do not systematically get either better or worse over the 12 rounds of play. So participants expected more competition in simple rounds, but more entered there anyway.

The second alternative explanation is that potential entrants systematically overestimated their own scores more on simple tasks than on difficult tasks. This explanation is also contradicted by the data—the opposite is actually true. Participants underestimated their scores on the simple quiz, reporting that they had gotten 4.41 out of 5 correct, when in fact they actually got 4.58. This difference is revealed to be significantly different by a comparison of actual versus self-reported scores in a  $(2) \times (4)$  within-subjects ANOVA performed at the level of the individual, where the four blocks served as the second within-subjects factor,  $F(1, 90) = 19.03, p < .001$ . On the difficult test, by contrast, participants overestimated their scores, reporting that they had gotten .95 correct when in fact they had only gotten an average of .41 correct,  $F(1, 90) = 78.20, p < .001$ .

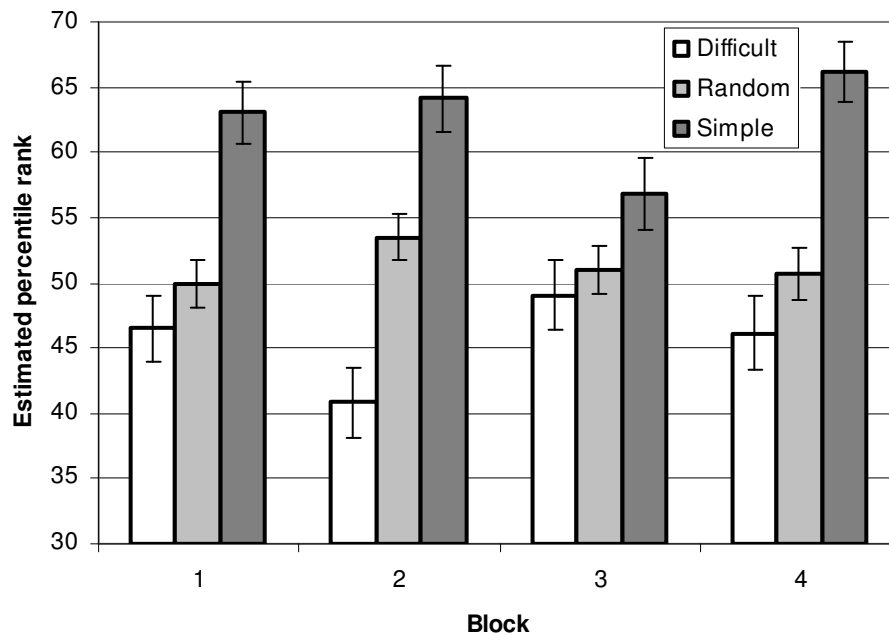
While this pattern at first seems incongruous, it ought not to be surprising: The tendency for people to overestimate their own performances more on difficult than on simple tasks is one of the more robust findings in the literature on overconfidence and calibration (Burson, Larrick, & Klayman, 2006; Lichtenstein, Fischhoff, & Phillips, 1982). It can be readily explained using the same regressive logic that we used to predict BTA effects on simple tasks and WTA effects on difficult tasks: Because people have imperfect knowledge of their own scores, their estimates of their own performances are slightly regressive (Erev, Wallsten, & Budescu, 1994; Juslin, Winman, & Olsson, 2000). If people's estimates of their own performances are slightly

regressive, then their estimates of the performances of others are likely to be even more regressive. This follows from the fact that people have better information about themselves than they do about others, and so, people underestimate others more on simple tasks than on difficult tasks.

In simple rounds, our participants underestimated their scores (which rules out the second alternative explanation) *and* they expected more competition—yet more entered there anyway. Before we test the reference group neglect explanation, let us turn to our explanation for the observed entry rates: differential regression. The data are consistent with differential regression. People underestimated others' scores on simple quizzes more than their own, reporting that others would score a regressive 4.2 out of 5, but that they themselves would score 4.41,  $F(1, 90) = 13.92, p < .001$ . On the other hand, participants overestimated others' scores on difficult quizzes more than their own, reporting that others would score a regressive 1.49 out of 5, but that they themselves would score .95,  $F(1, 90) = 41.63, p < .001$ .

Because participants' estimates of others are so regressive, they believe themselves to be above median on the simple quiz and below median on the difficult quiz. On the simple test, participants reported that they expected to outscore 63% of the other participants taking the same test. On the difficult test, by contrast, participants only expected to outscore only 46% of the others. On the simple test, participants expected there to be 5.2 entrants and expected that their rank among entrants would be 2.6. On the difficult test, participants expected only 3.2 entrants yet expected to rank 2.7. Figure 2 shows patterns in participants' beliefs about percentile rankings across the three treatments and four blocks. Differences between simple and difficult treatments persist throughout the experiment, despite the provision of feedback. There is little evidence for learning in this figure.

**Figure 2.** Participants' estimated percentile rankings (percentage of others worse than them) in the three different ranking systems across the four blocks. Error bars show standard errors.



These results provide a hint as to the reasons for the durability of differences in entry rates across different ranking systems. Participants got consistent feedback showing that they tended to underestimate their relative performances on the difficult quizzes and that they tended to overestimate their relative performances on the simple quizzes; but they nevertheless had specific new information about each quiz that might have undermined their willingness to attend to this general historical fact. Even if an individual notices that she has consistently overestimated her relative performance on simple quizzes, if she takes a new quiz and scores highly relative to her prior expectations, the inference that she is likely to be above average may still be a sensible one. So long as there is more uncertainty about others' scores than about her own, her predictions of others' scores will remain more regressive.

*Fourth explanation: Reference group neglect.* We have presented evidence for the idea that estimates of others are regressive. The remaining question is whether the differential regression explanation alone can account for the observed differences in entry rates between



experimental treatments, or whether there is any evidence of reference group neglect. Reference group neglect posits that participants chose to enter on simple rounds and stay out on difficult rounds not because they actually believed that they would score any differently from others—but because they just weren't paying enough attention to others (Klar & Giladi, 1997; Kruger, 1999; Windschitl et al., 2003). If people neglect to consider the group and instead focus on themselves when estimating their relative standing, we should observe that beliefs about own performance are weighted more heavily than are beliefs about others. In what follows, we test this prediction in a pair of regression analyses (in Table 2) predicting comparative judgments using performance by self and others. As the remainder of the results section details, the substantial majority of the effect of difficulty can be accounted for by greater regressiveness in estimates of others than of self.

Table 2. Actual and perceived value of participants' own scores and the scores of others for estimating percentile rank.

Model 1 predicting actual percentile rank			Model 2 predicting self-reported percentile rank		
Independent Variable	<i>B</i>	Std. Error	Independent Variable	<i>B</i>	Std. Error
Own actual score	.201*	.005	Own estimated score	.111*	.006
Actual average score	-.201*	.006	Estimated average score	-.077*	.008
<i>R</i> <sup>2</sup>	.56*		<i>R</i> <sup>2</sup>	.27*	

\**p* < .001

Model 1 in Table 2 is the optimal model, predicting participants' actual percentile ranks within each round using their own scores and actual average scores for the round as independent variables. The results show, not surprisingly, that participants' own actual scores and average

scores have B coefficients that are of similar size but opposite signs. We compare this result with participants' self-reported beliefs, using participants' beliefs about their own scores and beliefs about the average score to predict their self-reported percentile rank. The first apparent difference between these two analyses is more noise in self-reported beliefs than in actual performance, as shown by the smaller value of  $R^2$ . It ought not to be a shock that people's estimates of their own scores and their percentile rankings are imperfect.

The second and more important difference is that participants' beliefs about their own scores were weighted more heavily than were their beliefs about others' scores. To be precise, the weight attached to other ( $|B| = .077$ ) is 69% the size of the weight attached to self ( $B = .111$ ), and this difference is statistically significant,  $t(1088) = 3.4, p < .001$ . This shows evidence of reference group neglect but raises the following question: What proportion of our primary result (the effect of difficulty on entry rates) can be accounted for by differential regression and how much can be accounted for by reference group neglect? In order to answer this question, we first begin by assessing the experimental treatment's effect on entry decisions. We did this by regressing entry rates on experimental treatment. The independent variable in this regression was equal to 1 for simple-rank rounds, 0 for random-rank rounds, and -1 for difficult-rank rounds. When we conduct this analysis at the level of the round, the  $R^2$  value of this regression shows that the experimental treatment accounts for 28% of the variation in entry rates across all rounds,  $F(1, 154) = 59.67, p < .001$ . However, more useful to our purposes is this analysis performed at the level of the individual. There are two major reasons to expect  $R^2$  to be lower in the regression conducted at the individual level: First, participants' entry decisions are partially driven by their actual *relative* performance, which is uncorrelated with the experimental treatment; second, idiosyncratic individual-level factors such as risk preferences affect entry decisions. At the individual level, since the dependent variable is dichotomous (entry or not), a

logistic regression is the more appropriate statistical test.<sup>2</sup> The Nagelkerke  $R^2$  value of this logistic regression reveals that the experimental treatment accounts for 7.9% of the variation in individual entry decisions, and is statistically significant,  $\chi^2(1) = 65.67, p < .001$ . What this means is that 7.9% is the total size of the effect of difficulty on entry, and we must now determine how much of it can be accounted for by differential regression and how much of it cannot be.

In order to assess the effect of differential regressiveness, we next regressed entry decisions on participants' beliefs about their relative performance, as measured by the difference between their estimated absolute scores for self and for others. Beliefs about relative performance account for 24.4% of the variation in entry rates,  $\chi^2(1) = 218.96, p < .0001$ . The mere fact that participants' beliefs about relative performance are predictive of their entry decisions is neither impressive nor interesting—it would be surprising if they were not. The interesting question is whether these beliefs about relative performance can account for the effect of the experimental manipulations on entry decisions. In order to test for such a mediation effect, we conducted a third regression that included both experimental treatment and participants' self-reports of relative performance. The resulting Nagelkerke  $R^2$  value indicates that these two variables, combined, account for 26.0% of the variation in entry decisions. The inclusion of experimental treatment provides only a 6.7% increase in variation explained (over the 24.4% using only the relative performance). However the  $B$  coefficient associated with experimental condition remains significant ( $B = .34, SE = .09, p < .001$ ). The significance of experimental treatment suggests that there is an effect of difficulty that is distinct from

---

<sup>2</sup> For the sake of simplicity, we present logistic regression analyses in which each subject in each round serves as the unit of analysis (91 subjects \* 12 rounds = 1092 observations). The results we present are not appreciably different when the same analyses are conducted using a hierarchical linear model which treats subjects as random effects and accounts for the fact that experimental treatments are nested within trial blocks which are in turn nested within experimental sessions.

participants' beliefs about their relative standing. Of the total 7.9% of variation in entry decisions accounted for by our experimental treatment, 1.6% (or 26% minus 24.4%) cannot be accounted for by participants' self-reported beliefs about their own performance relative to others. This 1.6% represents 20% of the variation due to the experimental treatment that remains unexplained. Reference group neglect is the most viable alternative explanation for this unexplained 20%, but the substantial majority of the effect of difficulty (80%) can be accounted for by greater regressiveness in estimates of others than of self.

### Discussion

The results of the first experiment show that confidence regarding one's competitive performance depends on the type of competition. Contrary to prior evidence (Camerer & Lovallo, 1999; Harris, 1996; Klein & Kunda, 1994), we show that controllable tasks do not necessarily elicit more overconfidence than chance tasks. In difficult-rank rounds, people avoided entering. People overestimated others' performances, leading them to stay out of the competition despite the fact that they accurately forecast few other entrants. Thus, skill-based tasks do not always elicit overconfidence and entry rates depend in part on how difficult potential entrants see the task.

Participants' prior expectations regarding difficulty play an important role in our theory, but the first experiment did not measure them. Experiment 2 was designed to address this shortcoming. Furthermore, our theory posits a fundamental role for information about performance—one's own and others'. Experiment 2 allows us to observe the effect of information on participants' beliefs as they learn first about their own performances and then about the performances of others.

## EXPERIMENT 2: THE EFFECT OF INFORMATION ON COMPARISONS

Because our differential regression explanation describes the mechanisms by which errors in entry occur, it also offers useful insights into which interventions might be useful for reducing errors and which interventions are unlikely to be effective. Experiment 2 tests these interventions. Participants were first told that they would be taking either a difficult or simple quiz and were then asked to predict the outcome (Time 1). After taking the quiz (Time 2), participants were invited to revise their answers to their prior estimates of absolute and relative scores. Our theory would predict that information about one's own performance provided at Time 2 would produce BTA on easy tasks and WTA on difficult tasks. Finally, participants were given full information about how others scored on the same quiz they took, and they were asked to report the same comparative judgments (Time 3). Our theory would not predict BTA and WTA effects at Time 3, in the presence of excellent information about others. Previous research has shown that information about others can reduce BTA effects (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995). Here, we test whether it can also reduce WTA effects.

## Method

*Participants.* We recruited 128 undergraduate students at Carnegie Mellon University by offering them a base payment of \$2 plus from \$0 to \$8 on top of that. Experimental sessions were advertised under the name "*Games of skill*" with the following description: "*Participants will be playing a game in which they can earn money. How much you get paid will depend on exactly how things come out.*"

*Design.* The experiment had a 2 (quiz difficulty: simple vs. difficult) x (3) (time of wager: before quiz vs. after quiz vs. after results) mixed design. Quiz difficulty was manipulated between subjects and time served as a within-subjects factor.

*Procedure.* Participants were each given \$4 and invited to bet as much as they wanted on winning a trivia competition against a randomly chosen opponent. Participants were truthfully told that their opponents' scores would be selected at random from a group of 144 students who had previously taken these same quizzes as a part of a different study (reported in Moore and Kim, 2003, Experiment 3). None of the 128 participants in the present study had participated in that prior study. The test would consist of 10 items plus an eleventh tiebreaker question that virtually eliminated the chance of a tied score. Winning participants would double the amount they bet; those who lost would keep only the un-wagered portion of their \$4. Note that the second and third time they bet, participants were told that the most recent bet would be the one that counted.

Participants in the simple quiz condition were told they would take a simple trivia quiz and shown the following example question and answer:

*What is the common name for the star inside our own solar system?*

*Answer: the Sun*

Participants in the difficult quiz condition were told they would be taking a difficult trivia quiz and shown the following example question and answer:

*What is the name of the closest star outside our solar system?*

*Answer: Proxima Centauri*

Participants were then asked how much they wanted to bet. After they bet, participants were given a questionnaire that asked:

(1) *"How many of the 10 questions do you think you will get right?"*

(2) *"How many of the 10 questions do you think your opponent will get right?"*

(3) *"What percentage of the group will have scores below yours? (If you expect your score will be the very best, then put 100. If you expect your score will be exactly in the middle, put 50. If you expect your score will be the lowest, put 0.)"*

Questions 1 and 2 were objective measures of absolute evaluation for self and for opponent. Question 3, like the bet, was a direct measure of beliefs about relative standing. After participants had answered all these questions, they were given an actual trivia test. The questions from the difficult and simple quizzes are listed in Appendix C. Participants were then told, *"Now that you have taken the quiz, you may choose to revise your answers to these questions. Please answer all the questions, whether or not you put the same answers as before."* Then participants were asked how much they wanted to bet and were asked the same list of questions again. These were their Time 2 responses.

After they had answered all the questions at Time 2, participants were then given truthful feedback about the scores of the previous test-takers from whose ranks their randomly selected opponent would be drawn. For example, those who took the simple quiz were informed that: *"The average score is 8.71 out of 10, with a standard deviation of 1.1."* Those who took the difficult quiz were told: *"The average score is 1.48 out of 10, with a standard deviation of 1.01."* Participants were also given a breakdown of the percentage of others who got each of the 11 possible scores (from 0 to 10) on the quiz. After they had a chance to review this information, participants were told, *"Now that you have seen how others did, you may choose to revise your answers to these questions. Please answer all the questions, whether or not you put the same answers as before."* Then participants were asked how much they wanted to bet and were asked the same list of three questions again. These were their Time 3 responses. The bet that was counted for computing payoffs was this third and final one.

Our differential regression explanation holds that BTA effects and WTA effects result when people have good information about themselves but lack information about others, such as at Time 2 after taking the quiz. At Time 3, after getting good information about others, these effects should go away. Time 1 beliefs are useful for assessing participants' priors, but are based on so little information that our theory would not make strong predictions regarding their beliefs. We shall test both our differential regression explanation and that of reference group neglect.

### Results and Discussion

*Manipulation check.* As expected, the simple quiz resulted in higher scores ( $M = 8.25$  out of 10,  $SD = 2.01$ ) than did the difficult quiz ( $M = 1.54$  out of 10,  $SD = 1.34$ ),  $F(1, 126) = 490.39$ ,  $p < .001$ ,  $\eta^2 = .80$ .

*Participants' predictions at Time 1.* At Time 1, participants who had seen only an easy sample question (and were about to take—but had not yet taken—the simple test), predicted that they would score 7.22 ( $SD = 1.57$ ) and that others would score 6.41 ( $SD = 1.79$ ) out of 10. Those who saw only a difficult sample question predicted that they would score 5.22 ( $SD = 1.90$ ) and that others would score 4.92 ( $SD = 1.65$ ). We analyzed these predictions using a 2 (difficulty)  $\times$  2 (target: self vs. other) mixed ANOVA. The results reveal a main effect of target,  $F(1, 124) = 11.97$ ,  $p = .001$ ,  $\eta^2 = .09$ , since people predicted that they would do better than would others. The differential regression explanation cannot account for this effect; the results suggest some basic amount of self-enhancement. The main effect of difficulty is, of course, also significant,  $F(1, 124) = 46.13$ ,  $p < .001$ ,  $\eta^2 = .27$ . The target  $\times$  difficulty interaction effect does not attain statistical significance,  $F(1, 124) = 2.77$ ,  $p = .099$ ,  $\eta^2 = .02$ .

*Tests of differential regression at Time 2.* At Time 2, the differential regression explanation would hypothesize that people predict better relative performance (BTA) on simple tasks and (WTA) worse relative performance on difficult tasks. This would manifest itself in a



significant interaction between difficulty (simple vs. difficult) and target (self vs. other). Indeed, when we subjected estimates of absolute performance to this  $2 \times (2)$  ANOVA, the difficulty  $\times$  target interaction emerges as significant,  $F(1, 120) = 20.77, p < .001, \eta^2 = .15$ .<sup>3</sup> At Time 2, participants reported believing that they scored better ( $M = 8.30, SD = 1.49$ ) than their opponents ( $M = 7.83, SD = 1.28$ ) on the simple quiz,  $t(61) = 2.94, p = .005, \eta^2 = .12$ . But they also reported believing that they scored worse ( $M = 2.39, SD = 1.31$ ) than their opponents ( $M = 3.30, SD = 1.61$ ) on the difficult quiz,  $t(59) = -3.48, p = .001, \eta^2 = .17$ .

Consistent with our theory, the increase in BTA and WTA effects from Time 1 to Time 2 is largely attributable to changes in beliefs about one's own score. On average, participants changed their estimates of their own scores by 2.34 points ( $SD = 1.77$ ). However, they only changed their estimates of others' scores by 1.92 points ( $SD = 1.76$ ), and this difference is significant by paired-samples  $t$ -test,  $t(123) = 2.34, p = .02$ . Furthermore, these changes mediate the difference on bets between difficulty conditions from Time 1 to Time 2. We included these two change measures in a regression predicting change in participants' bets from Time 1 to Time 2, along with a dummy variable for difficulty. Their inclusion renders difficulty non-significant,  $\beta = .02, p = .87$ . As our theory would predict, changes in self-estimates were a significant predictor of changes in bets,  $\beta = .47, p < .001$ . However, changes in other-estimates were not significant,  $\beta = -.13, p = .16$ .

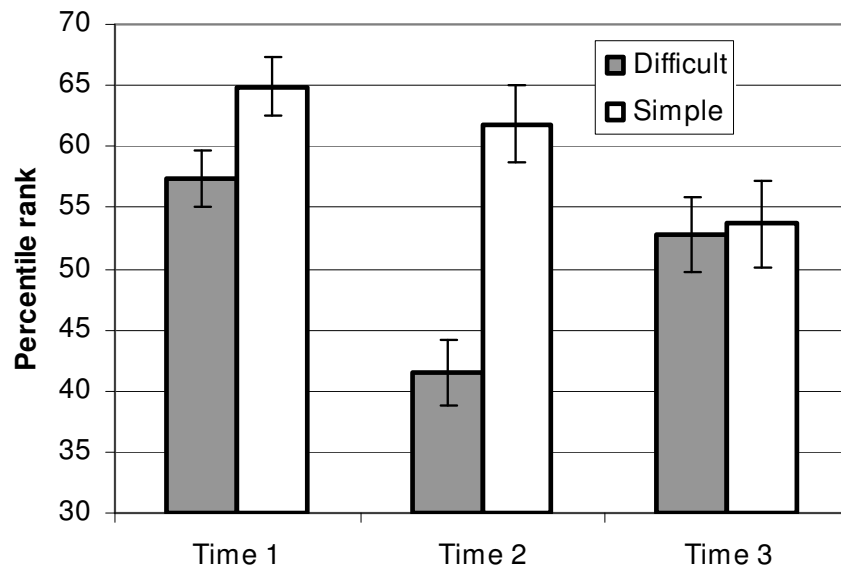
The differential regression explanation would not predict BTA and WTA effects at Time 3, when participants have good information not only about themselves but about others. Indeed, the same  $2 \times (2)$  ANOVA on absolute evaluations at Time 3 does not produce a significant target

---

<sup>3</sup> Naturally, the main effect of difficulty is also significant, since participants predict higher scores on the simple than on the difficult test,  $F(1, 120) = 606.16, p < .001, \eta^2 = .84$ . The main within-subjects effect of target (self vs. opponent) is not significant,  $F(1, 120) = 1.40, p = .24, \eta^2 = .01$ .

x difficulty interaction,  $F(1, 124) = .02, p = .89$ .<sup>4</sup> On the simple quiz, participants predicted similar scores for themselves ( $M = 8.31, SD = 1.62$ ) and for their opponents ( $M = 8.14, SD = 1.37$ ). Likewise on the difficult quiz, participants predicted similar scores for themselves ( $M = 2.08, SD = 1.19$ ) and for their opponents ( $M = 1.90, SD = .71$ ). Figure 3 shows participants' self-reported percentile ranks. Furthermore, as we would have predicted, estimates of others' scores changed more from Time 2 to Time 3 than did estimates of self. Estimates of others changed by an average of 1 point ( $SD = 1.18$ ), whereas estimates of self only changed by .4 points ( $SD = 1.05$ ), and these two are significantly different from one another  $t(123) = -4.57, p < .001$ . And consistent with our theory, the reduction in BTA and WTA effects on bets is mediated by changes in people's beliefs about others,  $\beta = -.31, p = .003$ , not the self,  $\beta = .16, p = .07$ .

**Figure 3.** Participants' estimated percentile rankings (percentage of others worse than them) in the two difficulty conditions at three points in time. Error bars show standard errors.



*Tests of reference group neglect.* The reference group neglect hypothesis predicts that direct comparisons (like estimates of percentile rank) will show stronger BTA and WTA effects than will indirect comparisons (computed by subtracting absolute estimates of others from self)

<sup>4</sup> Naturally, the main effect of difficulty remains significant,  $F(1, 124) = 931.98, p < .001, \eta^2 = .88$ . The main effect of target remains insignificant,  $F(1, 124) = 2.42, p = .07, \eta^2 = .03$ .

which make others' performance salient. The standard test is to regress comparative judgment on absolute evaluations of target and referent. Using this standard test, we replicate the result that comparative evaluation is strongly associated with self-evaluation but more weakly predicted by absolute evaluation of others. We regressed percentile estimates on predictions of point scores by self and other for responses at Time 1, before participants had taken the actual test. As Table 3 shows, the  $\beta$  coefficient for absolute self-evaluation is .86,  $p < .001$ , indicating that absolute and relative self-assessment are strongly correlated. The  $\beta$  coefficient for other-evaluation, however, is -.53,  $p < .001$ , is only 62% the magnitude of the coefficient for self. This finding is consistent with reference group neglect.

**Table 3.** Experiment 2's results for the three different measures of comparative judgment at three points in time. The third column shows the effect size of the difference between simple and difficult conditions, and asterisks the significance of the t-test comparing difficulty conditions. Regression results predicting indirect comparative judgment for the three different measures of comparative judgment appear in the fourth and fifth columns.

Time	Self-Reported Comparative Judgment	Simple vs. Difficult	Regression results	
		Effect size ( $\eta^2$ )	$\beta$ (Self)	$\beta$ (Other)
1	Bet	.02	.53***	-.27*
1	Percentile rank	.04*	.86***	-.53***
1	Indirect comparison	.03	1.07†	-1.0†
2	Bet	.15***	.92***	-.48**
2	Percentile rank	.16***	.93***	-.49***
2	Indirect comparison	.15***	1.86†	-1.53†
3	Bet	.03*	1.67***	-1.35***
3	Percentile rank	< .001	1.84***	-1.69***
3	Indirect comparison	.00	3.10†	-3.01†

\*  $p < .05$ , \*\*\*  $p < .001$ , † Independent variables perfectly account for dependent variable

Note that this differential weighting changes as people gain information. At Time 2, when participants had more information about themselves, the weight put on other-estimates (.49) is only 53% the size of the weight put on self-estimates. But at Time 3, when people had better information about others, other-estimates ( $\beta = 1.69$ ) carry 92% the weight placed on self-

estimates ( $\beta = 1.84$ ). If reference group neglect affects how people bet, then we ought to observe some effect of test difficulty on bets, over and above the effect of differential regressiveness on estimates of self and others' actual performances. We tested this as we did in Experiment 1.<sup>5</sup> The result was that 74% of the effect of difficulty on bets at Time 2 could be explained by differential regression. However, this test may claim too much credit for differential regression. As the results in Table 3 highlight, better information about self than others appears to produce both differential regression and differential weighting. When they are confounded, this test will give all the credit to differential regression over differential weighting.

At Time 3, the effect of test difficulty on bets shrinks dramatically: Difficulty accounts for only 3.4% of the variance in bets. Our theory would not predict that test difficulty would affect comparative judgments when people possess complete information regarding performance by self and others. Indeed, statistically speaking, differential regressiveness accounts for only 3% of this small effect. The remaining 97% is most likely attributable to the egocentric overweighting of self-knowledge over other-knowledge.

## GENERAL DISCUSSION

The present results offer two primary findings. First, the results of the first experiment show that confidence regarding one's competitive performance depends on the ease of the task. Contrary to prior evidence (Camerer & Lovallo, 1999; Harris, 1996; Klein & Kunda, 1994), we show that controllable tasks do not necessarily elicit more overconfidence than chance tasks. People underestimated others' performances on simple tasks but overestimated them on difficult tasks, leading them to enter with confidence on simple rounds despite the fact that they

---

<sup>5</sup> First, we began with the primary effect of test difficulty on bets. At Time 1, test difficulty only accounted for a statistically insignificant 1.9% of the variance in bets, as shown by the  $R^2$  value associated with a regression predicting bets using a dummy variable for experimental condition. At Time 2, however, difficulty accounts for 15% of the variance in bets.

accurately forecast numerous other entrants. Yet in difficult-rank rounds, people decided not to enter: They overestimated others' performances, leading them to stay out of the competition despite the fact that they accurately forecast few other entrants. Thus, skill based tasks do not always elicit overconfidence; instead, confidence and entry rates depend in part on how difficult potential entrants see the task.

The second contribution of this paper is that we identify the cause for what appear to be myopic interpersonal comparisons, namely, better information about one's own performance than about the performance of others. When a task is simpler than people expect it to be, people's estimates of others' performances regress downward and a majority will conclude that they are better than others. When a task is more difficult than expected, estimates of others' performances will regress upward and a majority will conclude that they are worse than others. Experiment 2 shows that these WTA and BTA effects are strongest when people are confident regarding their own performances but unsure of the performances of others. This may also explain why BTA and WTA effects are stronger when people compare themselves to some vague group than when they compare themselves to a specific, known individual (see Hoorens & Buunk, 1993; Klar, Medding, & Sarel, 1996; Klein & Weinstein, 1997; Perloff & Fetzer, 1986; Price, 2001; Windschitl et al., 2003). It may also help explain why BTA effects have been shown to be stronger for observable performances than for tasks or traits where people only know about themselves and cannot observe others' performances directly (Dunning, Meyerowitz, & Holzberg, 1989; Suls, Lemos, & Stewart, 2002). Given that differential regression explains most of the effect of difficulty on entry and on bets in our experiments, providing decision-makers with better information about the performance of others is likely to be the most effective way to eliminate this cause of myopic comparisons.

Our studies experimentally manipulated the information people had, making it impossible for us to measure differences in the degree to which people seek out information about themselves and others (for other studies that have measured this, see Moore, Oesch, & Zietsma, in press; Radzevick & Moore, 2006). However, our results nevertheless show that people do not always use information as they should. When making social comparisons, participants ought to have given the same weight to information about themselves as to information about others. The fact that one's competitors are doing poorly on some task is just as important as the fact that one is doing poorly. Consistent with the reference group neglect explanation, however, our participants' self-reported beliefs regarding their own performance were weighted more heavily than their beliefs regarding opponents' performance. But this egocentrism effect is *not* the driving factor behind WTA and BTA effects and their consequences for behavior. Differential weighting (reference group neglect) accounts for a small proportion of the effect of difficulty on entry rates and bets.

We should also note that, besides reference group neglect, there are other viable explanations for why people's judgments would appear to weight self more heavily than others. For example, people easily conflate relative and absolute evaluation on vague subjective measures (Baron, 1997; Biernat, Manis, & Kobrynowicz, 1997). The conflation error occurs when people answer the question, "How good are you relative to others?" as if they were answering the question "How good are you?" (Klar & Giladi, 1999). Unlike differential weighting due to reference group neglect, the conflation explanation is exceedingly mundane: Vague measures facilitate confusion between relative and absolute evaluation (Burson & Klayman, 2005; Moore, 2005). Conflation can make comparative evaluations appear to overweight the self because the person does not take herself to be making a comparative evaluation.

Although we have attempted to distinguish the differential regression explanation from reference group neglect, we should also note their fundamental compatibility. Both are caused by the greater accessibility and quality of information about the self than about others. Less information about others leads people to make more regressive estimates of others but also probably leads to further underweight of those regressive estimates (Kruger, Windschitl, Burrus, Fessel, & Chambers, 2006).

*Managerial and Economic Implications.* Our results have implications for understanding entrepreneurial entry. While we do not take a stand on the question of whether overall rates of actual entrepreneurial entry are excessive, we instead note that rates of entry vary considerably between industries. Some industries, such as retail clothing stores, restaurants, and bars, are marked by persistently high rates of entry and high rates of subsequent exit (U.S. Small Business Administration, 2003). Indeed, one of the stylized facts to emerge is that rates of entry and exit are highly and positively correlated (Dunne, Roberts, & Samuelson, 1988; Geroski, 1996; Mata & Portugal, 1994). Differences in rates of entry are not well accounted for by the size of an industry, the profitability of its firms, or barriers to entry (Geroski, 1996). So, if new firms enter an industry because of above-normal profits and exit an industry because of below-normal profits, then one would instead expect entry and exit to be negatively correlated such that high failure rates meant low entry rates, at least in the short term. Yet the correlation between entry and exit in any given year is around .7. The results presented in this paper suggest a possible explanation. Perhaps industries that see persistent high rates of entry are those that potential entrants view as "easy" or in which they feel capable (Greico & Hogarth, 2004). Such industries are then likely to see more intense competition, lower profits, and higher rates of failure. Experiment 1 shows that this might occur in spite of entrants' correct prediction that they will

have lots of competition in "simple" markets. Indeed, even when they *underestimate* their own absolute abilities, entrants will underestimate the abilities of their competition even more.

However, given that our experimental participants were not actual entrepreneurs with substantial quantities of money at stake, we must be cautious about generalizing our results from errors made in the lab. Might actual entrepreneurs learn to avoid the biases in comparative judgment shown by participants in the present experiments? While it is possible that experience may allow actual entrepreneurs to learn to overcome these errors, it is unclear how much experience is needed for such learning to take place. Participants in the market entry game were students at a selective university and also experienced 12 rounds with full feedback. Perhaps this experiment included too few trials for them to learn to solve this problem. If this is the case, however, entrepreneurs are likely to make the same mistake. After all, even the most experienced entrepreneurs rarely get the opportunity to start more than a handful of firms. Furthermore, our experimental setup was more transparent, assessment of the competition was clearer, and the causes of success were more obvious than they are likely to be for most entrepreneurs.

We must note that both our theory and our results are bounded. We do not claim that the differential regression explanation (a cognitive explanation) accounts for *all* BTA and WTA effects. Motivation and bias do influence comparative interpersonal judgments in important ways (Kunda, 1990). It is also clear that other cognitive explanations such as reference group neglect can account for some biases in comparative judgment and in strategic decision making, as shown in the present findings as well as in other work (Klar, 2002; Moore, 2004; Rose & Windschitl, 2006; Windschitl et al., 2003). However, the errors showcased in this paper seem to be more about not having good information about one's competition, as opposed to merely ignoring the competition.



### **Appendix A: Formalization of the Differential Regression Explanation**

In this appendix, we attempt to formalize our theory, first in general terms, then with a specific example. Our point in this appendix is not to suggest that it is possible to predict with certainty what individuals will believe—given a multitude of constraints that may or may not be realistic (we let our data speak to how real people behave)—nor do we intend to provide a generalized proof or justification of the decision process we are suggesting is at work. Here, we wish only to provide insight on why WTA/BTA effects might occur in groups of reasonable people.

Let us restrict our analysis to some specific group of  $n$  individuals who all expect to take a test together. These people, on average, expect that their performances will be average, relative to the others in the group. It is not necessary that everyone believe that they are exactly average, just that beliefs are balanced within the group. In other words, for every person (or subset of people) who believes that they are better than others (or better than some individual) there is another person (or subset of people) who believes that they are worse than others (or worse than some individual) to the same degree, and *vice versa*.

Within the group, let  $S$  be the average of all individuals' estimates of the absolute performance of *self*; let  $O$  = the average of all individuals' estimates of the absolute performance of *others*. To be precise:

$$S = \frac{(s_1 + s_2 + \dots + s_n)}{n}$$

where  $s_n$  is the  $n^{\text{th}}$  person's self-estimate.

$$O = \frac{(o_1 + o_2 + \dots + o_n)}{n}$$

where  $o_n$  is the  $n^{\text{th}}$  person's estimate of (the average of) others' scores (not including self).

$$S - O = \frac{(s_1 + s_2 + \dots + s_n) - (o_1 + o_2 + \dots + o_n)}{n}$$

and

$$S - O = \frac{(s_1 - o_1 + s_2 - o_2 + \dots + s_n - o_n)}{n}$$

$S - O$  is equal to the average comparative judgment of self to others. When  $S - O$  is positive, it implies a BTA effect: On average, people believe that they are better than others. When  $S - O$  is negative, it implies a WTA effect: People believe they are worse than others. At the outset, assume that  $S = O$  (on average, people expect that their performance will be average, relative to the others).

Let people acquire additional information about their own (past or anticipated) performance. To the extent that this information about self is different than prior information, it will justify updating beliefs about one's own performance (away from the prior). To the extent that information about the self is more useful for estimating performance by self than others, it will lead to greater updating of  $S$  than of  $O$ . Therefore, estimates of others will tend to be closer to the prior than will estimates of self. In other words,  $O$  regresses to the prior more than does  $S$ . We call this rule "Rule O":

Rule O: Since (whenever possible)  $O$  is more regressive than  $S$ ,  $O$  must be closer to the *prior* than is  $S$ . (When  $S$  is maximally regressive,  $O$  is equally regressive.)

There are three possible relationships between  $S$  and the *prior*:

Case 1:  $S > \text{prior}$ , meaning that, on average, people think that they did better than they expected to do; thus  $S > O$  (or else a contradiction follows; if we assume  $S > O$  to be false, i.e., if  $O \Rightarrow S$ , and (as given)  $S > \text{prior}$ , then  $O \Rightarrow S > \text{prior}$ , and  $O$  would be farther from the prior than  $S$ , violating Rule O), and people will, on average, believe they performed better than others.

Case 2:  $S < \text{prior}$ , meaning that, on average, people think that they did worse than they expected to do; thus  $S < O$  (ELSE:  $O \leq S < \text{prior}$ , and  $O$  would be farther from the prior than  $S$ , violating Rule O), and people will, on average, believe they performed worse than others.

Case 3:  $S = \text{prior}$ , meaning that, on average, people think that they did as they expected to do; thus  $S = O$  (ELSE:  $O > S = \text{prior}$  (or  $O < S = \text{prior}$ ), and  $O$  would be less regressive than  $S$ , violating Rule O), and people will, on average, believe they performed the same as others.

*Example:* Suppose there are three test takers, A, B, C, each completing a test that is scored out of 100. Suppose that, prior to taking the test, the average expected score is 50. Suppose A's actual score = 90; B's actual score = 65; C's actual score = 40. The average actual test-score = 65. *On average*, the test takers did better than expected ( $65 > 50$ ), even though some did worse than expected (e.g., C scored 40). Suppose, for the sake of simplicity, that all three know exactly how well they themselves did, but they know their sense of others is imperfect. Granted, specific numerical examples of estimates will depend on individual test-takers and specific tasks. With imperfect information, however, as in Bayesian updating, people's best estimates (in this case of others) will tend to fall between actual values and their prior expectations. Table 4 depicts

reasonable estimates. In keeping with the idea that estimates of self are less regressive than estimates of others, suppose that estimates of others are computed using a somewhat arbitrary equal weighting of the prior and actual score:  $(\text{prior} + \text{actual})/2$ . Since the average expected score is 50, assume that priors of all people's scores by all others = 50. So, for example, if A scores 90, A will predict that B scores  $(90 + 50)/2 = 70$ .

Table 4

(Row A, Col B) = B's prediction of A's score

		Predictor		
		A	B	C
Target	A	<b>90</b>	70	70
	B	57.5	<b>65</b>	57.5
	C	45	45	<b>40</b>

Next we calculate  $S$  and  $O$  for the group:

$$S = (\text{A's estimate of self} + \text{B's estimate of self} + \text{C's estimate of self})/3$$

$$= (90 + 65 + 40)/3$$

$$= 65$$

$$O = [\text{A's (average) estimate of others} + \text{B's (average) estimate of others} + \text{C's (average) estimate of others}]/3$$

$$= [1/2 (57.5 + 45) + 1/2 (70 + 45) + 1/2 (70 + 57.5)]/3$$

$$= [51.25 + 57.5 + 63.75]/3$$

$$= 57.5$$

Result:  $S (65) > O (57.5)$ . On average, relying on sensible rules of inference, but using systematically imperfect information (and which is known to be imperfect), people believe that they ( $S$ ) are better than others ( $O$ ).

Note that  $C$  actually does (40) worse than expected (50) and everyone knows it [but  $C$  knows it best; as shown in the preceding table, where  $(C, C) = 40$ , while  $(C, A) = 45$ , and  $(C, B) = 45$ ]. Nevertheless, on average, the group thinks it did better than expected (actual = 65 = estimated  $>$  expected = 50). Our theory holds that, when this occurs, people will (on average) think they did better than average. The  $S - O$  calculation bears this out:  $S - O = 7.5 > 0$ .

The logic outlined above works just the same if each person's estimate of his or her own score is also imperfect and known to be imperfect, and therefore it also regresses toward the prior. The only key requirement is that estimates of others be more regressive than estimates of self, and Rule O holds.

**Appendix B: Trivia Quizzes (four Simple, four Difficult), Experiment 1**

Simple Test 1

1. Who was the first president of the United States?
  2. How many inches are there in a foot?
  3. What does MTV stand for?
  4. On what continent is the country of Egypt?
  5. What is the most widely spoken language in the US, after English?
- Tiebreaker: What is the height of the Eiffel Tower (in feet)?

Simple Test 2

1. What was the first name of the Carnegie who founded the Carnegie Institute of Technology?
  2. How many states are there in the United States?
  3. In which month is Thanksgiving celebrated in the United States?
  4. Harrisburg is the capital of what US state?
  5. On what continent is the country of France located?
- Tiebreaker: How many films did Alfred Hitchcock direct?

Simple Test 3

1. Which American civil rights leader gave a famous speech in which he repeated the lines, "I have a dream..."
  2. What American director was behind the movies, A.I., E.T., Minority Report, Saving Private Ryan, and Jurassic Park?
  3. What is the name of Pittsburgh's professional hockey team?
  4. What Pennsylvania city is know for being at the confluence of the Allegheny and Monongahela Rivers?
  5. What country lies directly north of the United States?
- Tiebreaker: How many member states are there in the United Nations?

Simple Test 4

1. What American became the first person to ever win the Tour de France 6 times?
  2. Paris is the capital of what country?
  3. In what large US city is the famous Times Square located?
  4. Where in the human body is the cerebellum located?
  5. What famous act of military aggression by Japan happened on Dec 7, 1941 that brought the United States into World War II?
- Tiebreaker: How many men signed the Declaration of Independence?

Difficult Test 1

1. In what European city would you find the famous Tivoli Gardens?
2. Truth or Consequences is a city in what U.S. state?
3. What company's research and development lab was once known as the "House of Magic?"
4. What is the largest moon of Saturn?
5. What African country lies directly south of Egypt?

Tiebreaker: In the 2000 U.S. Census, what was the population of Walla Walla, Washington?

Difficult Test 2

1. Thomas Hooker is associated with the founding of which of the thirteen American colonies?
2. Who is the only U.S. president to have served two non-consecutive terms in office?
3. In Quentin Tarrantino's Reservoir Dogs, what is the alias of the man who is revealed to be an undercover police officer?
4. In The Odyssey, who was the son of Ulysses (Odysseus)?
5. Who was voted Time magazine's Man of the Year in 1938?

Tiebreaker: What is the land area of Morocco (in square kilometers)?

Difficult Test 3

1. Blues musician Huddie Ledbetter is better known by what name?
2. What make and model of car holds the record for being the most widely produced car in the world?
3. Laudanum is a form of what drug?
4. Who was the president of Indonesia, as of August 2002?
5. What is the capital of Nepal?

Tiebreaker: Approximately how many pieces of art did Pablo Picasso create during his lifetime?

Difficult Test 4

1. Which team won the first NBA Draft Lottery?
2. The Nobel Prizes are awarded in what two cities?
3. Dr. Faustus is best known for selling what item?
4. What two South American countries are land-locked?
5. Pro football announcer John Madden coached which team to a Super Bowl victory?

Tiebreaker: How many consecutive weeks did the Pink Floyd album Dark Side of the Moon spend on the billboard music charts?

**Appendix C**

Trivia questions used in the simple and difficult trivia quizzes (Experiment 2).

Simple	Difficult
1. How many inches are there in a foot?	Which creature has the largest eyes in the world?
2. What is the name of Pittsburgh's professional hockey team?	How many verses are there in the Greek national anthem?
3. Which species of whale grows the largest?	What company produced the first color television sold to the public?
4. Who is the president of the United States?	How many bathrooms are there in the White House (the residence of the U.S. President)?
5. Harrisburg is the capital of what U.S. state?	Which monarch ruled Great Britain the longest?
6. What was the first name of the Carnegie who founded the Carnegie Institute of Technology?	The word "planet" comes from the Greek word meaning what?
7. How many states are there in the United States?	What is the name of the traditional currency of Italy (before the Euro)?
8. What continent is Afghanistan in?	What is Avogadro's number?
9. What country occupies an entire continent?	Who played Dorothy in "The Wizard of Oz"?
10. Paris is the capital of what country?	Who wrote the musical "The Yeoman of the Guard"?

Tiebreaker question: How many people live in Pennsylvania?

Answers—Simple: (1) 12 (2) Penguins (3) Blue (4) George W. Bush (5) Pennsylvania (6) Andrew (7) 50 (8) Asia (9) Australia (10) France. Difficult: (1) Giant squid (2) 158 (3) RCA (4) 32 (5) Queen Victoria (6) wanderer (7) Lira (8)  $6.02 \times 10^{23}$  (9) Judy Garland (10) Gilbert & Sullivan. Tiebreaker: 12,281,054.



## References

- Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, 68(5), 804-825.
- Baron, J. (1997). Confusion of relative and absolute risk in valuation. *Journal of Risk and Uncertainty*, 14, 301-309.
- Biernat, M., Manis, M., & Kobrynowicz, D. (1997). Simultaneous assimilation and contrast effects in judgments of self and others. *Journal of Personality and Social Psychology*, 73(2), 254-269.
- 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.
- Camerer, C. F., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1), 306-318.
- 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).
- College Board. (1976-1977). *Student descriptive questionnaire*. Princeton, NJ: Educational Testing Service.
- Cooper, A. C., Woo, C. Y., & Dunkelberg, W. C. (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing*, 3(2), 97-109.

- Dunne, T., Roberts, M. J., & Samuelson, L. (1988). Patterns of firm entry and exit in U.S. manufacturing industries. *Rand Journal of Economics*, 19(4), 495-515.
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and business. *Psychological Science in the Public Interest*, 5(3), 69-106.
- 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.
- 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.
- Fischhoff, B., & Bruine De Bruin, W. (1999). Fifty-fifty = 50%? *Journal of Behavioral Decision Making*, 12(2), 149-163.
- Fox, C. R., & Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psychological Science*, 14(3), 195-200.
- Geroski, P. A. (1996). What do we know about entry? *International Journal of Industrial Organization*, 13(4), 421-441.
- Greico, D., & Hogarth, R. M. (2004). *Excess entry, ambiguity seeking, and competence: An experimental investigation*. Unpublished manuscript.
- 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.
- Hayward, M. L. A., & Hambrick, D. C. (1997). Explaining the premiums paid for large acquisitions: Evidence of CEO hubris. *Administrative Science Quarterly*, 42, 103-127.
- Hoelzl, E., & Rustichini, A. (2005). Overconfident: Do you put your money on it? *Economic Journal*, 115(503), 305-318.

- Hoorens, V., & Buunk, B. P. (1993). Social comparison of health risks: Locus of control, the person-positivity bias, and unrealistic optimism. *Journal of Applied Social Psychology*, 23(4), 291-302.
- 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.
- Klar, Y. (2002). Way beyond compare: Nonselective superiority and inferiority biases in judging randomly assigned group members relative to their peers. *Journal of Experimental Social Psychology*, 38(4), 331-351.
- Klar, Y., & Giladi, E. E. (1997). No one in my group can be below the group's average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology*, 73(5), 885-901.
- Klar, Y., & Giladi, E. E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin*, 25(5), 585-594.
- Klar, Y., Medding, A., & Sarel, D. (1996). Nonunique invulnerability: Singular versus distributional probabilities and unrealistic optimism in comparative risk judgments. *Organizational Behavior and Human Decision Processes*, 67(2), 229-245.
- Klein, W. M. P., & Kunda, Z. (1994). Exaggerated self-assessments and the preference for controllable risks. *Organizational Behavior and Human Decision Processes*, 59(3), 410-427.
- Klein, W. M. P., & Weinstein, N. D. (1997). Social comparison and unrealistic optimism about personal risk. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 25-61; , 1997 ix, 1450).

- 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., Windschitl, P. D., Burrus, J., Fessel, F., & Chambers, J. R. (2006). *The rational side of egocentrism in social comparisons*. Unpublished manuscript.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498.
- 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. (2004). *CEO overconfidence and corporate investment*. Cambridge, MA.
- Mata, J., & Portugal, P. (1994). Life duration of new firms. *Journal of Industrial Economics*, 42(3), 227-246.
- Moore, D. A. (2004). 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. (2005). *When good = better than average*. Pittsburgh: Tepper Working Paper 2004-E38.
- 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., Oesch, J. M., & Zietsma, C. (in press). What competition? Myopic self-focus in market entry decisions. *Organization Science*.
- Myers, D. G. (1998). *Social psychology* (5th ed.). New York: McGraw-Hill.
- Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *Journal of Finance*, 53(6), 1887-1934.
- Perloff, L. S., & Fetzer, B. K. (1986). Self-other judgments and perceived vulnerability to victimization. *Journal of Personality and Social Psychology*, 50(3), 502-510.
- Price, P. C. (2001). A group size effect on personal risk judgments. *Memory and Cognition*, 29, 578-586.
- Radzevick, J. R., & Moore, D. A. (2006). *For the love of the game? Betting, prediction, and myopic bias in athletic competition*. Pittsburgh: Tepper Working Paper 2005-E7.
- Rose, J. P., & Windschitl, P. D. (2006). *How egocentric optimism change in response to feedback in repeated competitions*. Unpublished manuscript.
- Suls, J. M., Lemos, K., & Stewart, H. L. (2002). Self-esteem, construal, and comparisons with the self, friends, and peers. *Journal of Personality and Social Psychology*, 82(2), 252-261.
- Svenson, O. (1981). Are we less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47, 143-151.
- U.S. Small Business Administration. (2003). *Longitudinal Establishment and Enterprise Microdata*. Washington, DC: Office of Advocacy (202-205-6530).
- Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5), 806-820.

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.

Zajac, E. J., & Bazerman, M. H. (1991). Blind spots in industry and competitor analysis:

Implications of interfirm (mis)perceptions for strategic decisions. *Academy of*

*Management Review*, 16(1), 37-56.