

Running head: COMPARATIVE JUDGMENT

Error and Bias in Comparative Social Judgment:
On Being Both Better and Worse Than We Think We Are

Don A. Moore
Carnegie Mellon University

Deborah A. Small
University of Pennsylvania

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

Phone: (412) 268-5968
Fax: (412) 268-7345
don.moore@alumni.carleton.edu

Key words: Focalism, egocentrism, social comparison, overconfidence

The paper benefited from constructive comments by John Chambers, Jason Dana, Robyn Dawes, Joachim Krueger, and Paul Windschitl. Thanks to Bryan Routledge and Peter Boatwright for helping us think like Bayesians. Thanks also to Joshua Furgeson, Young Lee, Sam Swift, Sapna Shah, and Wemi Peters for help with data collection. The authors appreciated the support of a Berkman Junior Faculty Development grant from Carnegie Mellon University. Please send correspondence to don.moore@alumni.carleton.edu.

Abstract

Previous attempts to explain better-than-average and worse-than-average effects in comparative judgment have invoked bias and motivation as causes. This paper develops a more parsimonious explanation, the regression account, which assumes that people typically have better information about themselves than they do about others. When one's own performance is exceptional (either good or bad), it is often reasonable to assume others' will be less so. Consequently, people estimate the performance of others as less extreme (more regressive) than their own. The result is that people believe they are above average on simple tasks and below average on difficult tasks. This effect is increased by obtaining feedback about one's own performance, increasing the natural discrepancy between knowledge of self and others. The effect is attenuated by obtaining feedback about the performance of others.

Error and Bias in Comparative Social Judgment:
On Being Both Better and Worse Than We Think We Are

How good are we at assessing our own abilities? Evidence suggests we are not particularly good, or at least not as good as we think we are. We routinely overestimate ourselves relative to others. People believe that they are fairer, luckier, more virtuous, better drivers, and better investors than their peers (Epley & Dunning, 2000; Messick, Bloom, Boldizar, & Samuelson, 1985; Moore, Kurtzberg, Fox, & Bazerman, 1999; Svenson, 1981; Weinstein, 1980). Yet recent evidence suggests that such shameless self-aggrandizement is not universal (Kruger, 1999; Moore & Kim, 2003; Windschitl, Kruger, & Simms, 2003). Indeed, people consistently rate themselves *below* average in some desirable domains. People report themselves to be below average in juggling, unicycle riding, their probability of living past 100, and their chances of finding a \$20 bill on the ground in the next two weeks (Chambers, Windschitl, & Suls, 2003; Kruger, 1999; Kruger & Burrus, 2004). The characteristic feature of such worse-than-average effects is that they occur in domains where success (in objective terms) is rare. Most people are not good at juggling and most people do not live past 100. By contrast, domains in which people tend to rate themselves above average are domains in which absolute performance is high or at least in which people generally feel capable.

Even when people are asked to make judgments that are explicitly comparative, their responses often appear to be myopic solo judgments of themselves. This “solo comparison” effect describes the tendency for people to rate themselves better than average (BTA) on tasks where they perform well and worse than average (WTA) on tasks where they perform poorly. This effect cannot be explained by simple lack of motivation to actually make comparative

judgments. Moore and Kim (2003) invited participants in their studies to bet money on whether they would beat a randomly selected opponent in a trivia contest. When the trivia test was easy, participants bet significantly more than when the test was difficult, even though they knew they would only be competing against others who had taken the same test.

Prior Explanations

Researchers have proposed both motivational and non-motivational accounts for BTA and WTA effects (see Chambers & Windschitl, 2004). While there are clear reasons why people may be motivated to view themselves in a positive light, motivational accounts generally have trouble explaining WTA effects. Two broad classes of non-motivational explanations have been proposed to explain coexisting BTA and WTA effects in comparative self-evaluation.

The first type of explanation highlights the individual-to-group comparison problem. According to this explanation, people have trouble comparing an individual to a group, and errors in this process lead to biases in comparative evaluation (see Giladi & Klar, 2002; Klar, 2002; Klar & Giladi, 1997). Research by Alicke et al. (1995) is consistent with this view. They found that the BTA effect was stronger when people were asked to compare themselves to a group than to an individual. While comparison to a group *does* appear to exacerbate the BTA and WTA effects, it *is not* a necessary precondition to obtain the effects. Windschitl, Kruger, and Simms (2003) obtained both BTA and WTA effects when people were comparing themselves with a single known individual. More importantly, this explanation is vague with respect to exactly how individual-to-group comparison problems produce BTA and WTA effects.

The second type of explanation is the focalism account. According to focalism explanations, people focus on a target individual and attend disproportionately to that

individual's performance. Egocentrism results when the self is the focus of attention, but similar effects have been obtained when people are led to focus on individuals other than the self (Klar, 2002; Moore & Kim, 2003; Suls, Lemos, & Stewart, 2002). Differential weighting has been offered as the cognitive mechanism at work behind focalism. Its basic contention is that the target of judgment (usually the self) is weighted more heavily than is the referent (usually other people). In its most extreme form, the differential weighting explanation holds that people don't actually perform a comparative judgment; when asked, "How happy are you compared to other people?" people instead answer the question, "How happy are you?" (Klar & Giladi, 1999). A less extreme version of the differential weighting argument is that the focal target exerts greater weight on the comparative judgment because it anchors the comparative judgment (Kruger, 1999), because it is more accessible in memory (Kuiper & Rogers, 1979; Perloff & Fetzer, 1986), or because it is more salient or individuated (Alicke et al., 1995).

The typical evidence for differential weighting comes in the form of regression results examining the degree to which beliefs about a target's relative performance are driven by estimates of absolute performance by the target of the judgment and the referent to which the target is being compared. Estimates of absolute performance by the target and the referent are the independent variables in this regression, and comparative judgments of the target serve as the dependent variable. The standard result is that absolute evaluations of the target are strongly correlated with comparative judgments but that absolute evaluations of the referent are not. This result has been interpreted as meaning that absolute evaluations of the target are overweighted in comparative judgment. There are problems with this interpretation. First, if people make sensible estimates of the referent group, then the variance in it should be smaller than the variance in estimates of the self. Indeed, if people estimated the group perfectly, then there

would be no variance in it and it would be worthless for predicting comparative judgments.

Second, comparative judgments are often elicited using vague measures that are prone to being conflated with absolute evaluations.

The Present Explanation

This paper proposes a new cognitive mechanism behind BTA and WTA effects that makes fewer assumptions than does the differential weighting explanation. The regression explanation offered here simply points out that people usually have more information about themselves than they do about others. As a result, people generally make more accurate estimates of their own performance than of the performance of others. Estimations of others' performances are therefore likely to be regressive: less extreme than estimations of one's own performance. BTA effects on simple tasks and WTA effects on difficult tasks naturally follow. To see why this is so, imagine that you have taken a test on which you knew all of the correct answers. How did you do relative to others? Unless you think that everyone is exactly like you, then your own score is imperfectly diagnostic of their scores. If you got a question right, others will get it right more likely than not. However, the probability they get it right is less than 100%. As such, your estimate of their performance should be less extreme than your own. You must be above average. The converse holds for tests so difficult that you knew none of the answers: Your own failure is imperfectly predictive of others' failures, so a sensible prediction of others' scores would be that they would score better than you. You must be below average.

The regression explanation depends fundamentally on the principle of social projection. People routinely use beliefs about themselves as a basis for their beliefs about others, mistakenly projecting their own personal perspectives and experiences on to others (Krueger, 2000, 2003; Krueger, Acevedo, & Robbins, in press; Ross, Greene, & House, 1977; Van Boven, Dunning, &

Loewenstein, 2000). People who are thirsty assume that others are also thirsty (Van Boven & Loewenstein, 2003), and people who volunteer for a cause assume others will do the same (Ross et al., 1977). Some of these “false consensus” effects may be rational, given that the self is relevant to anticipating what the average person will do (Dawes & Mulford, 1996). However, Krueger and Clement (1994) show that the effects are stronger than rational inference would predict. Although there may be disagreement about its normative status, all of these studies concur in showing some social projection that is less than complete. People assume that others are like them, but less so.

BTA and WTA effects would disappear if everyone assumed that others were *identical* to themselves. However, research has demonstrated the many ways in which people assume themselves to be different from others. Some of these so-called “false uniqueness” effects may be driven by people’s desire to view themselves as more virtuous (Epley & Dunning, 2000) or principled (Allison, Messick, & Goethals, 1989; McCall & Mattrass, 2001) than others. On the other hand, it may often be reasonable for people who know that they are particularly good at something to assume that others might not be quite as exceptional. Conversely, those who know they engage in some behavior only rarely might reasonably assume that others are likely to do it more often than they do. McFarland and Miller (1990) demonstrated both false consensus and false uniqueness in the same study on different dependent variables. Moore and Kim (2003) demonstrated both false consensus and false uniqueness on the same dependent measure: Participants’ own scores on a trivia test were highly predictive of their beliefs regarding others’ scores. At the same time, people believed that their own scores would be more exceptional than they actually were: Those who took the simple test believed that they had scored better than others, while those who took the difficult test believed they had scored worse than others.

Formalizing the Argument

In the section that follows, we have formalized the regression account. For the sake of exposition, let us begin by considering beliefs about performance on a test with just one problem, where the answer can be either right or wrong. Before having seen the problem, and without any information regarding its ease or difficulty, one may begin by assuming that performance will be evenly distributed across possible outcomes (Fischhoff & Bruine De Bruin, 1999; Fox & Rottenstreich, 2003). This assumption translates to a prior belief that 50% of people will get the problem wrong and 50% will get the problem right. Such a uniform prior might make sense in the absence of better information.

Let's imagine that I take the test. Imagine that I have solved the problem and I know that I have solved it. To determine how my own performance compares with that of another person, I must estimate how likely it is that the other will be able to solve that problem. Whether or not I have solved it is, of course, useful for updating my beliefs about the difficulty of the problem and another person's probability of solving it. The solution to this problem is provided by Bayes's Theorem, which states that:

$$(1) \quad p(H | E) = \frac{p(E | H) p(H)}{p(E)}$$

Where: H is the hypothesis that another person got the problem right

E is the evidence that I got the problem right

$p(H|E)$ is the probability that the other got the problem right, given that I did

$p(H)$ is my prior belief about the probability that other would get the problem right

$p(E)$ is my prior belief about the probability that I would get the problem right

$p(E|H)$ is my belief in the probability that I got the problem right, given that the other did

This last number is crucial. If knowing whether someone else has solved the problem is useful for gauging its difficulty, then given that they have solved the problem, the probability that I have solve the problem should be higher than the .5 ignorance prior. If there are just two types of problems, simple ones which everyone gets right and difficult ones which everyone gets wrong, then $p(E|H)$ is equal to 1. The more the other person's score is useful for predicting my own score, the farther $p(E|H)$ is from the ignorance prior and the closer it is to 1.

A uniform prior means that, in the absence of better information, $p(H)$ and $p(E)$ are both equal to .5. Filling this in to the equation leaves us with:

$$(2) \quad p(H | E) = \frac{p(E | H) (.5)}{(.5)}$$

These numbers cancel on top and bottom, leaving us with:

$$(3) \quad p(H | E) = p(E | H)$$

In other words, Bayesian logic dictates that the best estimate of whether someone else has gotten a given problem right is the degree to which knowing that the other person has gotten the problem right is useful for predicting whether I will get it right. This, of course, is a direct function of the correlation between the scores of different people. This correlation is naturally high, but it is less than 1 unless everyone scores identically. If $p(H|E)$ is less than 1 and I got the problem right, it is sensible for me to infer that I am probably above average. After all, I know with certainty that I got the problem right. I cannot be as sure that others will also get the problem right. As a result, my estimates of others' performances will tend to regress to the ignorance prior. If I perform well, I will predict that others will do worse than I did. If I perform poorly, I will predict that others will perform better than I did.

It is simple to extend this logic to tests with more than one item. If we assume for the sake of simplicity that all the problems on the test are of equal difficulty, then the predicted score

for others is simply the number of correct answers multiplied by the probability that others got an answer right, given that I did, added to the number of incorrect answers multiplied by the probability that others got an answer wrong, given that I did. For example, imagine that I am trying to determine whether I did better or worse than my friend on a 10-item test on which I know I got 8 questions right. I assume my friend will perform very similarly to me. For each item I got right, I expect there is a 90% chance my friend also got it right. For each item I got wrong, I expect there is a 10% chance my friend got it right. To compute my friend's score, I would simply multiply the number I got right by .9, and add this to the number I got wrong multiplied by .1. This leads me to expect that my friend will score 7.4.

The formalization we have offered here is based on Bayes's rule. People do not automatically use Bayes's Rule when updating prior probabilities based on evidence (Dawes, 1998; Kahneman & Tversky, 1972; Tversky & Koehler, 1994), but research demonstrating people's violations of Bayes do not contradict the basic prediction we are making. The most consistent violations of Bayes's Rule follow from the representativeness heuristic, in which people do not appropriately weight the prior hypothesis, such as by ignoring base rates (Kahneman & Tversky, 1973). The most extreme form of this argument would hold that people ignore the prior completely. In the present case, that would lead people to assume that others' scores were exactly the same as their own. However, while studies tend to show some base-rate neglect, they do not usually ignore the prior completely (see El-Gamal & Grether, 1995). As long as people do not predict that others will be identical to themselves, their predictions of others will be regressive. In other words, when solving a problem, people do not assume that everyone else will necessarily also solve it, and that if they fail to solve it, they do not assume that no one else will be able to solve it.

Naturally, the extent to which people's predictions of others regress from their own performances to the ignorance prior depends on how similar they believe they are to the person whose score they are trying to predict. If I believe that I am completely different from the person whom I am trying to predict, then my own score is irrelevant, and the knowledge that I got a problem right should not shift my prediction about their performance off of the 50% prior. However, if I believe that I am similar to the person whose behavior I am trying to predict (such as if I am trying to predict my own performance on a future task), then my own performance is very useful indeed.

How much do people believe they are like others? This is a question whose answer depends greatly on the situational specifics (Mabe & West, 1982). One relevant empirical estimate can be gleaned by re-examining data from Moore and Kim's (2003) third and fourth experiments. In these two studies, participants took 10-item trivia quizzes just as participants did in the first experiment we present here. They were then asked to predict their own scores and the scores of other people. The results clearly show that they thought their own scores were relevant to predicting the behavior of others. The correlation between people's estimates of their own scores and their estimates of the scores their randomly selected opponents was .89. Another way to examine the relationship between people's estimates of themselves and of others is to conduct a regression predicting people's estimates of their (randomly selected) opponent's score using their estimates of their own scores. We found that the unstandardized B coefficient was equal to .76, meaning that for every correct answer, participants' predictions of others' scores went up by an average of .76.

Thus far, our analysis has assumed that people's estimates of their own scores were perfectly accurate. Of course, after having just attempted to solve the problem, people can make

good estimates of the likelihood that they actually solved it. Depending on the type of problem, they may know for certain whether they have solved the problem, or they may still be uncertain. We again turn to the results from Moore and Kim's (2003) Experiments 3 and 4 for an empirical estimate of the accuracy of self-evaluation on the sorts of trivia-test tasks presented in the first experiment in this paper. In Moore and Kim's data set, the correlation between actual score on the 10-item trivia quiz and participants' estimates of their performance was .91. When actual scores are used as independent variables in a regression predicting people's estimates of their own scores, the resulting B coefficient is .78.

These numbers imply a graph with three lines: one for their actual scores, one for their estimates of their own scores, and one for their estimates of others' scores. That graph is shown in Figure 1. This is a stylized picture of the consequences of the process we have described. People begin with a sense of the difficulty of the test and first use that to make an inference about what their score is likely to be, but these inferences are imperfect, and as a result people tend to underestimate their own performances on simple tasks and overestimate them on difficult tasks. People then use their beliefs about their own performances to make inferences about the performances of others, but they do not believe that others are identical to themselves. The result is that estimates of others tend to be less extreme than are their predictions of themselves, and people end up believing that they must be above average on simple tasks and below average on difficult tasks.

There are a few noteworthy features of Figure 1. First, the true mean scores lie along the diagonal, and the deviations from that line represent error. Second, WTA effects are present when absolute scores are low and BTA effects are present when absolute scores are high, consistent with prior evidence of biases in comparative judgment. Third, people underestimate

their own scores on simple tasks and underestimate them on difficult tasks. This feature of the regression explanation reconciles findings of BTA and WTA effects in comparative judgment with the so-called hard/easy effect in overconfidence.

The hard/easy effect in overconfidence describes the fact that people overestimate their own absolute performances more on difficult tasks than on easy tasks (Erev, Wallsten, & Budescu, 1994; Lichtenstein & Fischhoff, 1977; Soll, 1996; see also Gigerenzer, Hoffrage, & Kleinboelting, 1991). The more difficult the task, the more people tend to overestimate how they did. Initially the hard/easy effect in overconfidence seems to contradict BTA and WTA effects, but in fact they are readily reconciled. The hard/easy effect pertains to judgments of absolute performance, whereas BTA and WTA effects pertain to comparative judgments. Both can exist simultaneously, as they do in Figure 1. With respect to estimates of absolute performance, people overestimate both their own *and* others' performances on difficult tasks and underestimate both on simple tasks. As long as people's estimates of others' performances are more regressive than their own, the result will be BTA effects on simple tasks and WTA effects on difficult tasks. Thus, people can be overconfident in absolute terms but underconfident in relative terms on difficult tasks, and the reverse on easy tasks.

We do not claim that the regression explanation accounts for *all* BTA and WTA effects. Motivation and bias do influence comparative judgments in important ways—indeed the results of the experiments we present here show that they do. Yet the simulated data in Figure 1 demonstrate that BTA and WTA effects need not entail motivation or bias beyond sensible Bayesian inference. In reality, the influence of bias in this domain is likely weaker than previously assumed by researchers, since a substantial proportion of BTA and WTA effects

result from the simple fact that people have more information with which to estimate themselves to estimate others.

The Present Research

In this paper, we examine some of the predictions made by the regression explanation. We present the results of three experiments. In the first experiment, we collect measures of actual performance, as well as estimates of performance by self and others. This allows us to draw a graph like Figure 1 using actual rather than simulated data. Experiment 1 also examines the degree to which people will project beliefs about their own performance on to others. The regression explanation as we have presented it applies to people comparing themselves with others. When people are comparing other individuals to each other and do not apply self-knowledge differentially to those other people, then the regression account would not predict biases in comparative judgments. In order to test this implication, Experiment 1 includes a condition in which participants are asked to compare two randomly chosen individuals. We find that BTA and WTA effects are weaker in this condition.

At first, the results of Experiment 1 appear to contradict Moore and Kim's (2003) results with respect to focusing on another person. The second experiment we present is a re-analysis of data from Moore and Kim's fourth experiment. Here we show how the discrepant findings can be reconciled.

The third experiment directly manipulates participants' knowledge about their own and others' performances. We show that, consistent with the regression explanation, providing better information about the self reduces the regressiveness of self-estimates, thereby reducing the hard/easy effect. At the same time, however, it increases BTA and WTA effects because estimates of others remain regressive. Providing better information about others' performances,

on the other hand, reduces BTA and WTA effects. In other words, BTA and WTA effects result from a natural discrepancy in information about the self and others. Increasing that gap by providing more information about the self enhances BTA and WTA effects, while narrowing the gap by providing more information about others reduces the discrepancy.

EXPERIMENT 1: THE TRIVIA QUIZ

Method

Participants. We recruited participants after classes at Carnegie Mellon University. An experimenter invited students in each of six classes to remain for 10 minutes and complete an experiment for cash payment. Two-hundred fifteen individuals agreed to participate.

Procedure. Each participant was given a packet of questionnaires, beginning with a 10-item trivia quiz and an 11th tiebreaker question (see Moore & Kim, 2003). Half the participants received a simple quiz including questions like “How many inches are there in a foot?” whereas others received a difficult quiz including questions like, “What is Avogadro’s number?” The tiebreaker question (“How many people live in Pennsylvania?”) was scored based on participants’ distance from the correct answer, and virtually eliminated the possibility of a tied score.

Next, instructions informed participants that they had earned \$3 and that they could wager any amount of it on the trivia competition. If they bet and won, the amount of the bet would be doubled. Half the participants bet on whether they would beat a randomly chosen opponent (the tiebreaker question resolved tied scores). The other half of participants bet on whether a randomly selected (anonymous) protagonist would beat a randomly chosen opponent. For those betting on this random protagonist, the random nature of the selection was driven home

by asking them to draw a number out of a hat in order to determine the protagonist on whom they would be betting.¹

The experiment, therefore, employed a 2 (difficulty) × 2 (protagonist) between-subjects design.

Dependent measures. Participants made both comparative and absolute judgments of performance. Comparative measures included participants' bets, estimates of their probability of winning, and their responses to the question "How do you expect that you [the person whose number you drew] will score relative to all other people taking the same test?" The response scale ran from 1 to 7, with labels at 1 (*well below average*), 4 (*average*), and 7 (*well above average*).

Participants' scores on the actual quiz served as measures of absolute performance. In addition, participants were asked to estimate scores for the protagonist (self or the randomly chosen protagonist) and the randomly chosen opponent. We assessed confidence in these estimates by asking participants to specify (for both the protagonist and opponent) scores above and below their guesses such that they were 90 percent sure the true score fell within that range.

Results and Discussion

Manipulation check. Scores on the simple test were in fact higher ($M = 8.51$ out of 10, $SD = 1.48$) than were scores on the difficult test ($M = 1.66$ out of 10, $SD = 1.26$), $t(213) = 36.54$, $p < .001$.

Comparative judgments. To assess the effects of the manipulations on comparative judgments, we subjected participants' bets to a 2 (difficulty) × 2 (protagonist) ANOVA. The results reveal a main effect for difficulty: Those taking the simple quiz bet more ($M = \$1.84$, $SD = \$1.01$) than did those taking the difficult quiz ($M = \$1.38$, $SD = \$1.14$), $F(1, 210) = 11.34$, $p <$

.01, $\eta^2 = .05$. These results parallel results of asking participants to compare the protagonist (either self or random person) with the group average on a 7-point scale. Again, the same 2×2 ANOVA reveals a main effect for difficulty, such that the protagonist is rated better than average in the simple quiz condition ($M = 4.55$, $SD = .97$) but worse than average in the difficult quiz condition ($M = 3.66$, $SD = .96$), $F(1, 200) = 64.12$, $p < .001$, $\eta^2 = .24$.² However, this main effect is qualified by a significant difficulty \times protagonist interaction, $F(1, 200) = 23.25$, $p < .001$, $\eta^2 = .10$. The effect of difficulty on comparative evaluation was greater for those betting on self than for those betting on a randomly selected person (see Table 1).

Absolute judgments. To test the effect of the experimental manipulations on participants' estimates of absolute performance for self and opponent, we conducted a 2 (difficulty) $\times 2$ (protagonist: self vs. randomly chosen person) $\times 2$ (target: protagonist vs. opponent) mixed ANOVA with repeated measures on target. Naturally, the results reveal a significant between-subjects main effect of difficulty: Participants taking the simple test predicted higher scores ($M = 7.5$, $SD = 1.4$) than did participants taking the difficult test ($M = 3.0$, $SD = 1.5$), $F(1, 173) = 443.36$, $p < .001$, $\eta^2 = .72$. This main effect is qualified by a significant target \times difficulty interaction, $F(1, 173) = 25.36$, $p < .001$, $\eta^2 = .13$. Participants estimated that the easy test would be easier for the protagonist ($M = 7.9$, $SD = 1.5$) than for the opponent ($M = 7.4$, $SD = 1.5$), $t(87) = 3.45$, $p = .001$, whereas the difficult test would be more difficult for the protagonist ($M = 2.9$, $SD = 1.6$) than for the opponent ($M = 3.3$, $SD = 1.5$), $t(88) = -2.49$, $p = .014$. This two-way interaction is qualified by the expected three-way target \times difficulty \times protagonist interaction, $F(2, 173) = 8.72$, $p < .01$, $\eta^2 = .05$. This three-way interaction reveals that, consistent with the regression explanation, differences between scores predicted for protagonist and opponent are greater for those betting on themselves than for those betting on a randomly selected person.

Among those taking the simple test and betting on themselves, the self was predicted to score higher ($M = 8.37$, $SD = 1.42$) than was the opponent ($M = 7.68$, $SD = 1.02$), $t(29) = 3.37$, $p = .02$. This difference did not attain significance among those betting on others, $t(57) = 1.84$, $p = .071$. Among those taking the difficult test and betting on themselves, the self was predicted to score worse ($M = 2.47$, $SD = 1.31$) than was the opponent ($M = 3.10$, $SD = 1.42$), $t(29) = -2.39$, $p = .023$. This difference was not significant among those betting on others, $t(58) = -1.00$, $p = .321$.

It is worth noting that the main within-subject effect of target is not significant, $F(1, 173) = .46$, ns , $\eta^2 = .003$. This is because people estimated that the person on whom they were betting would score about as well, on average ($M = 5.40$, $SD = 2.95$), as the opponent ($M = 5.33$, $SD = 2.57$). Perhaps more strikingly, this held true even among those betting on themselves. There is no evidence of overall self-enhancement: People estimated that they would score about as well ($M = 5.33$, $SD = 3.39$) as would their randomly selected opponents ($M = 5.34$, $SD = 2.63$).

Measures of regressiveness. Participants were imperfect estimators of performance. As shown in Figure 2, for those betting on themselves, estimates look regressive; people's estimates are less extreme than their actual scores.³ Consistent with the assumption that individuals have more information about themselves than about others, participants estimated their own scores with greater accuracy than others' scores. Nevertheless, people estimated their own scores with some error. When performance was low, they overestimated it; those taking the difficult test only got an average of 1.66 ($SD = 1.26$) correct, yet estimated that they got 2.22 ($SD = 1.34$) correct, $t(105) = 4.45$, $p < .001$. When performance was high, they underestimated it; those taking the simple test got an average of 8.51 ($SD = 1.59$) correct, yet estimated that they got 8.37 ($SD = 1.46$) correct, $t(106) = -7.28$, $p < .001$. This effect shows up as a significant difficulty x

measure interaction, $F(1, 202) = 66.6, p < .001, \eta^2 = .25$, in a 2 (difficulty) \times 2 (measure: actual vs. estimated score) ANOVA with repeated measures on the second factor. A regression predicting participants' estimates of their own scores using their own actual scores as the independent variable reveals that getting one more item right increased participants' predictions of their own scores by .87 of a point.

Each participant estimated a 90 percent confidence interval for both protagonist and opponent. Participants knew that their estimates of their own scores were not perfectly accurate; they established confidence intervals around their answers that were, on average, 2.97 points in width ($SD = 1.40$). Their confidence intervals are even wider when estimating others' performances ($M = 4.16, SD = 1.97$), $t(60) = -5.47, p < .001$. In a 2 (difficulty) \times 2 (protagonist) \times 2 (target: protagonist vs. opponent) mixed ANOVA with repeated measures on target, the main between-subjects effect of protagonist, $F(1, 173) = 11.45, p = .001, \eta^2 = .06$, and within-subject effect of target, $F(1, 173) = 29.91, p < .001, \eta^2 = .15$, are significant, but they are qualified by a significant target \times protagonist interaction, $F(1, 173) = 37.75, p < .001, \eta^2 = .18$, reflecting greater confidence when predicting their own scores than when predicting others' scores (see Table 1).

We also expected that participants would use the self as a basis for estimating others, which would also contribute to greater regressiveness in estimates of others. In other words, for those betting on themselves, actual performance would serve as the basis for predicting the performance of the protagonist, which would in turn serve as the basis for predicting the performance of the opponent. In order to examine this hypothesis, we conducted a mediational analysis (Baron & Kenny, 1986) in which actual performance and self estimate were used to predict opponent estimate. The results appear in Figure 3A, and indicate that estimates of self

mediate the effect of actual performance on estimates of opponent. The converse, however, is not true (see Figure 3B); opponent estimates do not fully account for the effect of test performance on estimates of self.

Although participants used their own experiences as guide for estimating the performance of others, they did not assume that others would be exactly like themselves. To quantify the degree to which participants expected other to be like them, we conducted another regression using participants' predictions of the scores of their (randomly selected) opponents as the dependent variable and participants' predictions of their own scores as the predictor variable. The results show that, among those who were betting on themselves, believing that they scored one more item right on the test increased participants' predictions for their opponents' scores by .72 of a point. The result is regressiveness in participants' estimates of others. After just having overestimated their own performances, those taking the difficult quiz overestimated the performances of others to an even greater extent. Similarly, those who had taken the simple quiz underestimated their performances, yet underestimated the performances of their opponents even more.

Mediational tests of comparative judgment. If it is the more regressive predictions of others that lead to both BTA and WTA effects, then the estimated differences between one's own score and those of others should mediate the effect of test difficulty on comparative judgments for those who were betting on themselves. In order to test this hypothesis, we conducted three sets of regressions, shown in Figure 3C. First, comparative judgments were regressed on a dummy variable for test difficulty. Indeed, those who took the difficult test rated themselves as worse than those who took the simple test. In the second regression, test difficulty had a dramatic influence on the predicted difference between scores for self and for opponent. When

both test difficulty and predicted score difference are used to predict comparative judgments, both remain significant. Differences between predicted absolute scores for self and the group account for 60 percent of the variance in comparative self-evaluation. Nevertheless, these predictions do not fully mediate the relationship between difficulty and relative evaluation because test difficulty remains a significant predictor of comparative evaluation after controlling for differences in predicted absolute scores.⁴

Summary. The results of Experiment 1 are consistent with the regression explanation. People used their own performances to estimate their own performances, but because they did not perfectly estimate how well they had done, their self-estimates tend to be regressive. As a result, consistent with the hard/easy effect on overconfidence, they estimate that they did worse than they actually did on the simple test and better than they actually did on the difficult test. But their estimates of others were even more regressive than were their estimates of themselves. As a result, they believed that they were better than average on simple tasks and worse than average on difficult tasks.

EXPERIMENT 2:

A REANALYSIS OF DATA FROM MOORE AND KIM (2003, EXPERIMENT 4)

There is a noteworthy contrast between the findings of the first experiment and the findings presented by Moore and Kim (2003). Moore and Kim argued that BTA on simple tasks and WTA on difficult tasks are not necessarily due to egocentric self-focus. They found that people displayed BTA and WTA effects when comparing two other people with each other if they were focusing on one of the two. The reduction in BTA and WTA effects among people betting on a randomly selected individual in Experiment 1 at first seems to contradict this

finding, but in fact is consistent both with a focusing account and with the present regression explanation. The key difference between the present Experiment 1 and Moore & Kim's experiment is that while Moore and Kim's participants were led to focus on the protagonist, the procedures of the present study impeded focusing on the protagonist. In the present study participants drew the protagonist's number from a hat, highlighting the random and anonymous nature of the protagonist. It is important to reiterate that the regression explanation is a mechanism that is intended to help explain focusing (and egocentrism) effects without reliance on the questionable assumptions inherent in the differential weighting explanation.

In their fourth experiment, Moore and Kim showed that focusing on the anonymous protagonist was key to obtaining BTA and WTA effects when betting on others. They were able to reverse the standard effect by getting their participants to focus on the opponent. In other words, when people were focusing on their opponents they predicted that their opponents would beat them (and their randomly-chosen protagonists) on simple tests and would lose to them on difficult tests. This reversal is driven by the fact that people actually believed that others would do better than they (or the protagonist) would on simple tests and worse than them on difficult tests. The results make it appear that when they were focusing on others, people's predictions of themselves are more regressive than are their predictions of others. This occurred despite the fact that people clearly had more information about themselves than they did about their opponents.

Here, we present a re-analysis of the Moore and Kim data that demonstrates why their findings do not, in fact, contradict the regression explanation presented here. The key to this reconciliation is the fact that information about the difficulty of some task that is gleaned through personal experience may be used as a basis for estimating the performance of others. It is our

contention that this is what happened in Moore and Kim's fourth experiment. They took the test and then applied this experience to estimating the performance the focal person. This score was then used as a basis for estimating the performance of the out-of-focus person. As a result, performance estimates of the out-of-focus individual appear to be more regressive than do estimates of in-focus individuals. We test this explanation in a re-analysis of the data from Moore and Kim's fourth experiment.

Method

Moore and Kim had 113 individuals take a 10-item trivia test that was either very simple (mean score = 9.47) or very difficult (mean score = 1.70). After they had taken the test, participants were told that they would have \$4 to bet the trivia quiz competition. Half the participants were told that they would be betting on whether a randomly selected protagonist would beat a randomly selected opponent. The other half of the participants were told that they would be betting on whether they themselves would beat a randomly selected opponent.

This protagonist manipulation was crossed with a focusing manipulation. The focusing manipulation was designed to induce participants to focus on one side of the competition or the other. Because self-focus is undoubtedly the norm, other studies that have employed mild focusing manipulation have achieved only modest reductions in the BTA and WTA effects (Eiser, Pahl, & Prins, 2001; Windschitl et al., 2003). Moore and Kim employed a more heavy-handed focusing manipulation that had three major elements. First, participants were given perspective-taking instructions that told them to put themselves in the perspective of the person on whom they were focusing. Second, participants read a brief autobiographical essay supposedly written by the person on whom they were focusing.⁵ Third, participants were asked questions that forced them to think about the in-focus competitor:

- 1) How many of the 10 trivia questions do you think [the target] answered correctly?
- 2) How do you expect that [the target] scored relative to all the other people taking the same test? (from 1-well below average to 7-well above average)
- 3) What percentage of the group will have scores below [the target]?
- 4) How likely do you think it is that [the target's] score will exceed that of [the referent]?

After answering these questions, participants were asked to make their bets. These bets served as the experiment's primary measure of comparative judgment.

Re-analysis and Discussion

Two key features of the results must be addressed here. First, those who were focusing on the protagonist bet more on the simple ($M = \$3.24$, $SD = \$1.09$) than the difficult test ($M = \$0.79$, $SD = \$0.90$), $t(56) = 9.32$, $p < .001$, and this is true both for people betting on themselves and people betting on an anonymous (and randomly selected) protagonist. Second, this effect reversed itself among those focused on the opponent. These participants bet more on the difficult ($M = \$2.52$, $SD = \$1.19$) than on the simple test ($M = \$1.18$, $SD = 1.22$), $t(53) = 4.13$, $p < .001$.

This result did not occur because people focused exclusively on the absolute performance of the protagonist and ignored his or her relative performance. As Moore and Kim's Table 4 shows (reproduced here as Table 2) people's responses suggest that they were weighting both the in-focus and the out-of-focus competitor. Instead, the BTA and WTA effects on bets occur because people's estimations of the out-of-focus competitor were more regressive than were their estimations of the in-focus competitor. People gleaned useful information about the test when they actually took it. After they took the test, they were asked about the in-focus competitor, and they used what they knew about the test to help them estimate the score of the in-focus

competitor. Then they went on to use this to estimate the performance of the out-of-focus competitor.

The result is that estimates of the in-focus competitor mediate the effect of an individual's actual test performance on their estimates of performance by the out-of-focus competitor. Figure 4 shows these mediation analyses. When people were focusing on the protagonist (either the self or the randomly chosen protagonist), then the effect of people's own actual performances on the trivia quiz, as measured by their scores, on estimates of the opponent's score, are perfectly mediated by estimates of the protagonist's score. However, when people were focusing on their opponents, then estimates of their opponents mediate the effect of their own scores on estimates of the protagonist's score.

The reconciliation between Experiment 1 and Moore & Kim's fourth experiment, then is two-fold. First, Moore and Kim used a much more powerful focusing manipulation, and it is this manipulation that produced significant BTA and WTA effects in the condition where people were betting on whether a randomly selected protagonist would beat a randomly selected opponent. Second, the focusing manipulation powerful enough that people used their experiences taking the test as a basis for estimating the in-focus competitor, which in turn was used as a basis for estimating the out-of-focus competitor. The result is that the performance of the out-of-focus competitor was estimated more regressively than was the performance of the in-focus competitor. We think it is worth pointing out that the result Moore and Kim's fourth experiment creates a rather unnatural situation. The self is almost always the focal actor in each individual's social world, and it is a rare situation when people are focusing on others rather than themselves. Experiment 3 returns the focus to the self and instead manipulates the quality of information participants have about themselves and about others.

EXPERIMENT 3:

HOW GOOD ARE YOU AT GUESSING HOW MUCH OTHER PEOPLE WEIGH?

The regression explanation posits a key role for the effect of information in comparative judgment. BTA and WTA effects should be magnified by high quality information about the one's own performance, but they should be reduced or eliminated by high quality information about others' performance. We test these predictions using a task which, unlike the tasks used in prior studies, provides participants with little sense of how they have done. In the past, comparative judgment has been studied using tasks for which participants have better information about themselves than they do about others. While this is realistic, it confounds the target of judgment and the quality of information about performance. It is theoretically possible to disentangle people's beliefs about themselves with the effect of information about performance by studying a task in which participants do not know how well they have done after they have completed it. Experiment 3 uses just such a task. This allows us to manipulate whether participants know how well they did or not. We also manipulate whether participants know how well others did. To do this, participants got feedback about their own performances, others' average performance, both, or neither.

Experiment 3 also attempts to improve measurement weaknesses of previous studies. A potential weakness of Experiment 1 and 2 is the measures of comparative judgment. A bet provides an incentive-compatible method for assessing people's beliefs that provides real rewards for accuracy. However, the way the bet is structured confounds comparative judgment with risk preferences. People could bet more either because they believe that they are better than others or because they are risk-seeking. The solution to this problem is to force people to bet all their money but give them two options. Bet 1 pays off if you are above average. Bet 2 pays off

if you are below average. This way, the amount of money that they bet on Bet 1 is a purer measure of participants' belief that they are above average.

Another measure of comparative judgment used in Experiment 1 and 2 was participants' responses to the question, "How do you expect that you will score relative to all other people taking the same test?" on a 7-point scale. While this sort of verbally-anchored subjective response scale is commonly used as a measure of comparative judgment, it does raise some concerns. After all, a variety of evidence suggests that people are prone to reinterpret the meaning of the scale's verbal anchors (Biernat & Manis, 1994). This can lead people to conflate relative with absolute evaluation (Baron, 1997; Moore, 2004). So in the third experiment, we use a new measure of comparative judgment that asks participants to make an explicit estimate of the difference in scores between themselves and the average.

An additional change in Experiment 3 is that participants compared themselves to a group average, whereas in Experiment 1 and 2 people were competing against a single randomly chosen opponent. When competing against a randomly selected opponent, participants might want to consider not only the mean, but also the variance and the skew in others' scores. Given evidence suggesting that BTA and WTA effects tend to get worse when the other side gets more complicated to consider (Alicke et al., 1995; Klar & Giladi, 1997; Windschitl et al., 2003), we elected this time to only have people comparing themselves with the group average.

Method

Participants. Three hundred and thirty five individuals participated in exchange for payment. Of these, 176 were students and staff at the University of Pennsylvania and 158 were students at Carnegie Mellon University. All were given \$4 to bet on their own performance and results of these bets yielded an average payment of \$4.92 ($SD = \2.84). The experiment lasted

approximately 15 minutes. Participants' ages ranged from 18 to 58, with a mean of 22.68 years ($SD = 7.25$ years). Fifty-six percent of participants were female.

Design. The experiment had a 2 (simple vs. difficult) \times 2 (feedback about self: yes vs. no) \times 2 (feedback about others: yes vs. no) between-subjects factorial design.

Procedure. After participants had arrived at the lab and signed consent forms, they were seated at computers and directed to a site on the world wide web that randomly assigned participants to conditions and led them through the experiment (<http://mktgweb.wharton.upenn.edu/deborahs/>). The instructions on the first page began: “*How good are you at figuring out how much other people weigh? In this exercise, you will be shown a series of pictures of other people and your task will be to guess, within 30 [4] pounds, how much they weigh. After you complete this weight-guessing test, you will be asked some questions about the test and your performance on it. You will also receive \$4 to bet on your performance in this exercise.*”

Participants the were shown a series of 10 photographs and had to guess how much each person in each picture weighed. Those in the simple condition got an item right when they were within 30 pounds of the true weight. Those in the difficult condition got an item right when they were within 4 pounds of the true weight. The 10 photographs appeared in a different random order for each participant.

After they had estimated the 10 weights, participants were given feedback about performance. In all conditions, feedback was truthful. Half the participants were informed of their own scores (out of 10) on the test. This manipulation of feedback about self was crossed with a manipulation of feedback about others. Half the participants were informed of the average score on that test, using the same criterion as they had (either 30 pounds or 4 pounds), of

330 people who had previously taken the test.⁶ This set of previous test-takers scored an average of 8.76 using the simple criterion and 1.27 using the difficult criterion.

After getting feedback, participants were asked to make estimates of relative and absolute performance. Half the participants answered the comparative performance question first and half answered the absolute performance questions first.

Comparative judgments. Participants were asked to compare their performances with the average performance using the following question: “*How many points above or below average do you think your score will be? Please estimate the difference in scores between yourself and the average score. Use **positive** numbers to indicate that you think you will be **above** average. Use **negative** numbers to indicate that you think you will be **below** average. For example, if you think that you got 1 more answer right than the average, put 1; if you think you got 2 more answers right than the average, put 2. If you think that you got 1 less answer right than the average, put -1; if you think you got 2 less answers right than the average put -2. If you expect to be exactly average, put 0.*”

Estimates of absolute performance. Participants were asked to make estimates of absolute performance by self and others with the questions: (1) “*How many of the 10 items do you think **you** got right?*” and (2) “*How many of the 10 items do you think **others** got right on average?*” The order in which these two absolute questions appeared was counterbalanced.

Bets. Participants were then asked how they wanted to bet their \$4. They had two bets on which they had to wager all their money: “*You have two bets, between which you must allocate your entire \$4. Money wagered on Bet 1 is doubled if your score is above average. Money wagered on Bet 2 will be doubled if your score is below average. Naturally, you may choose to put \$2 on each of the two bets, in which case you are guaranteed to make \$4, because one of the*

two bets will be guaranteed to win. If you are sure you are above average, you should bet all \$4 on Bet 1, because then you could make \$8. Similarly, if you are sure you are below average, you should bet all \$4 on Bet 2. In the unlikely event that your score is exactly equal to the average of all scores, the outcome of the bets will be determined randomly.”

After they had placed their bets, participants made a comparative judgment like the one made by participants in Experiment 1. They were asked to rate their performances relative to others on a 7-point scale that ran from 1 (*much worse*) to 7 (*much better*). Participants were also asked, “*What proportion of your performance on the weight-guessing task do you believe was caused by your own personal skills and abilities at guessing weights? Please answer using a percentage between 0% and 100%.*”

Results and Discussion

Comparative judgments. We tested the effect of the experimental manipulations on comparative judgments using a 2 (difficult vs. simple) \times 2 (self feedback: yes vs. no) \times 2 (other feedback: yes vs. no) ANOVA in which participants’ estimates of the differences between their scores and the average score served as the dependent variable. The results reveal main effects for difficulty and for self-feedback, as well as two-way interactions for difficulty \times self feedback and difficulty \times other feedback.⁷

The main effect for difficulty replicates BTA and WTA effects found previously. Participants in the simple condition tended to rate themselves as being more above average ($M = .80$, $SE = .15$) than did participants in the difficult condition ($M = -.17$, $SE = .16$), $F(1, 325) = 18.72$, $p < .001$, $\eta^2 = .05$. The main effect for self feedback reflects the fact that participants who did not get feedback about their own performances tended to believe they were more above

average ($M = .75$, $SE = .16$) than did participants who did get feedback ($M = -.13$, $SE = .16$), $F(1, 325) = 15.47$, $p < .001$, $\eta^2 = .05$.

Influence of feedback about self. The main effects reported above are qualified by a significant difficulty \times self feedback interaction, $F(5, 320) = 5.28$, $p < .001$, $\eta^2 = .08$. It is illustrated in Figure 5, and shows that when people lacked good information about themselves, their comparative self-evaluations were persistently over-optimistic, but there was no significant difference between the simple ($M = .84$, $SD = 1.91$) and difficult conditions ($M = .74$, $SD = 2.24$) with respect to how optimistic they were, $t(167) = .31$, $p = .76$. However, consistent with the regression explanation, good information about their own performances produced differences between the difficult and simple conditions. When people got clear information about their own performances, they reported themselves to be above average in the simple condition ($M = .76$, $SD = 1.96$) and below average in the difficult condition ($M = -1.01$, $SD = 2.06$), $t(162) = 5.61$, $p < .001$.

Influence of feedback about others. The second of the two-way interaction is the difficulty \times other feedback interaction, $F(1, 325) = 8.74$, $p < .001$, $\eta^2 = .03$. It is illustrated in Figure 6, which shows that, consistent with the regression explanation, clear information about others' performances rendered differences between simple and difficult conditions nonsignificant, $t(172) = .54$, $p = .59$. However, when participants lacked clear information about others, they reported themselves to be below average in the difficult condition ($M = -.52$, $SD = 2.61$) and above average in the simple condition ($M = 1.10$, $SD = 1.94$), and this difference is statistically significant, $t(157) = 4.50$, $p < .001$.

If regression were the only cause for BTA and WTA effects, then we would predict that the standard BTA and WTA effects would reverse themselves when people *only* had information

about others and not about themselves. In other words, if people's estimates of their own performances are regressive, but their estimates of others' performances are accurate, then they should predict that they would be worse than others at simple tasks and better than others at difficult tasks. The results for this condition do show a reversal of the standard effect: Those in the difficult condition reported themselves to be more above average ($M = 1.04$, $SD = 1.45$) than did those in the simple condition ($M = .66$, $SD = 1.93$), but this difference is not statistically significant, $t(89) = 1.06$, $p = .292$.

This calculation, however, ignores the fact that this small difference between simple and difficult conditions represents a significant reversal from participants' baseline beliefs (in the absence of any information) in the no feedback condition. In the no feedback condition, those in the simple condition estimated that they were (nonsignificantly) more above average ($M = 1.00$, $SD = 1.90$) than did those in the difficult condition ($M = .31$, $SD = 2.99$), $t(76) = 1.24$, $p = .219$.⁸ Given that this represents participants' prior beliefs, we can ask how information about others' performances led people to update their beliefs. To do this, we computed the difference in comparative evaluations between the no feedback condition and the other-feedback-only condition. The result is that, consistent with the regression hypothesis, information about others' performances (in the absence of information about themselves) led people to believe that they themselves were below average in the simple condition ($M = -.34$, $SD = 1.93$) and above average in the difficult condition ($M = .73$, $SD = 1.45$), and this difference is statistically significant, $t(89) = 3.01$, $p = .003$.

Mediational analyses. While the results in Figures 5 and 6 are consistent with the regression explanation, the regression explanation makes a more specific prediction. In particular, that BTA effects on simple tasks and WTA effects on difficult tasks will be driven by

differences in underlying beliefs about absolute performances by self and others. In other words, the effect of test difficulty should be reduced to insignificance after controlling for estimates of the absolute performance of self and others. In order to test this prediction, we began with a regression predicting estimated difference between self and average using the difficulty manipulation. Test difficulty is, as reported previously, a significant predictor of comparative judgment, $B = -.20$, $t(331) = -3.64$, $p < .001$.

However, when estimates of absolute performance by self and others are included as independent variables in the regression, the predictive value of test difficulty drops to insignificance, $\beta = .05$, $t(329) = .83$, $p = .41$, despite the high overall significance of the regression, $R^2 = .55$, $F(1, 329) = 136.50$, $p < .001$. This result shows that comparative judgments were driven by absolute evaluations of self and others. Absolute evaluations of self receive a higher beta-weight in this regression analysis, $\beta = 1.26$, $t(329) = 18.87$, $p < .001$, than do absolute evaluations of others. $\beta = -.96$, $t(329) = -14.72$, $p < .001$. The weight placed on others is only 76% the size of that placed on self. While this difference is consistent with the differential weighting explanation, the highly significant weighting on absolute evaluations of others is a marked contrast to findings in which it held no statistically significant relationship with comparative judgments (Klar & Giladi, 1997; Kruger, 1999; Windschitl et al., 2003). It would be a mistake to attribute the difference in beta weights to participants putting greater weight on self than others when making comparative judgments, because this difference may be attributable to the fact that there is simply less variance in people's estimates of others ($SD = 2.63$) than there is in estimates of themselves ($SD = 3.17$). Levene's test for the equality of variances is statistically significant, $F(1, 668) = 19.99$, $p < .001$, allowing us to reject the null hypothesis that the two variances are the same.

One meaningful comparison would be to ask how these weightings would compare if participants were making their comparative judgments in an unbiased manner, given their absolute evaluations. We can test this by conducting another regression analysis using the same independent variables but replacing the dependent variable (participants' comparative judgments) with the computed comparison implicit in their absolute evaluations of self and others (the "indirect" comparative judgment; Chambers & Windschitl, 2004). This analysis reveals that the different variances in evaluations of self and others do lead to differences in weightings, even if participants had made their comparative judgments in an entirely unbiased way. The beta-weight on self-evaluations is 1.67, while the beta-weight on evaluations of others is only 83% as large, $\beta = -1.38$.⁹ See Table 3. The results in Table 3 highlight the fact that different measures vary in their susceptibility to WTA and BTA effects. Note that the indirect comparison (the difference implicit in participants' absolute estimates of performance by self and others) is about as good as a direct measure of the difference in performance by self and others. This questions the generality of the conclusion that BTA and WTA effects tend to be stronger on direct than on indirect measures of comparative judgment (Chambers & Windschitl, 2004). This conclusion was likely based on comparing indirect measures with vague direct measures. The subjectively verbally-anchored response scale shows stronger BTA and WTA effects than do the indirect comparisons.

BTA and WTA effects are stronger when questions are more vague. It is noteworthy that bets are not better measures of comparative judgments. It is intuitively compelling to assume that incentives for accuracy should lead people to put forth greater effort. Behavioral researchers in the economic tradition insist that decisions made in the face of real consequences are both more interesting and less biased (Friedman, 1998; Smith & Walker, 1993). However, accuracy

was not greater in the incentive compatible bet than in the comparative judgment that lacked monetary incentives. This suggests again that BTA and WTA biases are not the result of a lack of motivation for accuracy.

It is also worth comparing effect sizes across the four different conditions. The results presented in Table 4 show that the effect of the difficulty manipulation is strongest in the condition in which people get self feedback only, consistent with the regression explanation for BTA and WTA effects.

Absolute over- and underconfidence. The regression explanation holds that poor information about performance leads to regressiveness in people's assessments of their own and others' performance. We first tested the hypothesis that this regressiveness led to the hard/easy effect: the tendency to underestimate performance on simple tasks and overestimate it on difficult ones. To do this, we calculated an index of absolute overconfidence by subtracting each participant's estimate of his or her own score from his or her actual score. This measure was then subjected to a 2 (difficulty) \times 2 (self-feedback) between-subjects ANOVA. The main between-subjects effect of difficulty is significant, $F(1, 331) = 127.07, p < .001, \eta^2 = .28$. People underestimated their performances on the simple test ($M = -.70, SD = 1.88$) and overestimated their performances on the difficult test ($M = 1.82, SD = 2.26$).¹⁰

This main effect is qualified by the expected difficulty \times self-feedback interaction, $F(1, 331) = 24.97, p < .001, \eta^2 = .07$. This interaction is due to the fact that the main effect of difficulty was dramatically stronger when participants lacked feedback about their own scores. In the absence of self-feedback, those taking the simple test *underestimated* their scores by an average of .9 of a point ($SD = 2.16$), whereas those taking the difficult test *overestimated* their scores by 2.62 points ($SD = 2.41$), $t(168) = -10.03, p < .001$. Those who got feedback about

their performances still tended to underestimate them on the simple test ($M = -.51, SD = 1.58$) and overestimate them on the difficult test ($M = .85, SD = 1.60$), $t(163) = -5.47, p < .001$, but self-feedback dramatically reduced this effect.

Relative over- and underconfidence. The regression hypothesis further predicts that people's predictions of others will be regressive when they lack good information about others. In order to test the hypothesis that good information about others would eliminate BTA effects on simple tasks and WTA effects on difficult tasks, we calculated an index of relative overconfidence by subtracting each participant's estimate of his or her performance relative to the average from his or her actual relative performance. This measure was then subjected to a 2 (difficulty) \times 2 (other-feedback) between-subjects ANOVA. As expected, the main between-subjects effect of difficulty is significant, $F(1, 329) = 30.81, p < .001, \eta^2 = .09$. Those taking the simple test believed their relative performances to be better than they actually were ($M = .90, SD = 1.86$) whereas those taking the difficult test believed their relative performances to be worse than they actually were ($M = -.31, SD = 2.35$).

However, this main effect is qualified by the expected difficulty \times other-feedback interaction, $F(1, 329) = 15.57, p < .001, \eta^2 = .05$. This interaction reflects the fact that the main effect of difficulty was dramatically stronger when participants lacked feedback about others' scores. In the absence of other-feedback, those taking the simple test *overestimated* their relative performances by an average of 1.29 points ($SD = 2.03$), whereas those taking the difficult test *underestimated* their relative performances by .87 points ($SD = 2.70$), $t(157) = 5.77, p < .001$. Those who got feedback about the average score did not estimate their relative performance differently in the simple ($M = .50, SD = 1.59$) and difficult ($M = .14, SD = 1.93$) conditions, $t(172) = 1.36, p = .176$.

We conducted an additional test of the regression explanation in order to examine one of its premises. The regression explanation holds that regressiveness in estimates of others' performances is reduced by the belief that everyone will perform similarly. Logically, if individual differences in skill do not influence performance, different people will score similarly, there should be little regressiveness in estimates of others, and BTA and WTA effects should be weak. On the other hand, if test performance is determined by the unique skills and abilities of the test-taker, then people should expect others to score differently from them, there should be regressiveness in estimates of others, and we should expect to observe BTA and WTA effects. We can test this hypothesis because we asked our participants to estimate the proportion of test performance attributable to their own skills and abilities. We can then correlate their answers to this question with their comparative judgments. In the standard condition where people got feedback about their own performances but did not get information about others' performances, this correlation is positive for those who took the simple test, $r = .46, p = .001$. Consistent with the regression explanation, BTA effects get stronger when people believe that their performance is due to their own idiosyncratic abilities. The same correlation is negative among those taking the difficult test, but it does not attain significance, $r = -.16, p = .37$. This reversal from the simple to the difficult condition is consistent with the regression explanation, because WTA effects should be stronger when people believe that performance was due to their abilities.

Self-enhancement. We can use participants' estimates of their own and other's absolute performance to test for self-enhancement effects. Experiments 1 and 2 did not find much overall evidence of self-enhancement, but all participants had a good sense of their own performance. This is also true in the current experiment among those who received feedback about their own performances. Those who got self feedback did not estimate their scores ($M = 5.67$ out of 10, SD

= 3.55) to be greater than others' scores ($M = 5.50$, $SD = 2.79$), $t(164) = 1.12$, $p = .263$.

However, when participants lacked good information about their own scores, they believed their scores to be greater ($M = 5.91$, $SD = 2.75$) than others' scores ($M = 5.44$, $SD = 2.48$), $t(169) = 3.41$, $p = .001$. This latter finding is consistent with work suggesting that ambiguity increases opportunities for motivated self-enhancement (Dunning, Meyerowitz, & Holzberg, 1989; Klein, 2001; Kunda, 1990).

And it is self-enhancement and not other-derogation that leads to this effect. Those who did not get feedback about their own scores estimated their scores to be higher ($M = 5.91$, $SD = 2.75$) than they actually were ($M = 4.99$, $SD = 3.81$), $t(169) = 4.17$, $p < .001$. It did not arise because people underestimated others' performance. Indeed, people also estimated the scores of others to be higher ($M = 5.44$, $SD = 2.48$) than they actually were ($M = 4.94$, $SD = 3.81$), $t(169) = 2.26$, $p = .025$, but less so than for themselves (see Figure 7). The finding that ambiguity lets people believe themselves to be better than average, and that this result arises from self-enhancement (rather than other-derogation) is consistent with the findings of Epley and Dunning (2000). They found that self-serving assessments were more the product of inflated self-evaluation than of underestimating others.

We should reconcile these findings with one of the premises behind the regression explanation: that people are better at estimating themselves than they are at estimating others. We believe that this is usually the case—people almost always have more information about themselves than they do about others. Yet this is not to deny that the power of motivated reasoning will exert its strongest force on self-assessments (see Bass & Yammarino, 1991; Risucci, Tortolani, & Ward, 1989). For instance, while it is true that people have better information than do others about their quality of their own intimate relationships, they also have

stronger motivational influences on their beliefs about the relationship's durability (MacDonald & Ross, 1999).

When their own performances are ambiguous, it is more likely that people's judgments will succumb to motivated biases that will lead to self-enhancement. Yet such motivated biases may be operating at the same time as the regression effect that is the subject of this paper. Indeed, a 2 (target: self vs. other) \times 2 (simple vs. difficult) mixed ANOVA with repeated measures on target (estimating self vs. estimating others) reveals both a main effect of target and a target \times difficulty interaction.¹¹ The main within-subjects effect of target results from the fact that people estimated their own scores ($M = 5.79$, $SD = 3.17$) to be higher than those of others ($M = 5.47$, $SD = 2.63$), $F(1, 333) = 9.22$, $p = .003$, $\eta^2 = .03$. The significant target \times difficulty interaction effect, $F(1, 333) = 19.53$, $p < .001$, $\eta^2 = .06$, arises because people believe themselves ($M = 8.06$, $SE = .147$) to be better than others ($M = 7.31$, $SE = .116$) on the simple test, $t(173) = 5.80$, $p < .001$. However, those who took the difficult test did not believe that they ($M = 3.36$, $SE = .154$) scored better than others ($M = 3.56$, $SE = .122$), $t(160) = -.89$, $p = .37$. In the simple condition, the regression effect is consistent with self-enhancement. In the difficult condition, the desire to see themselves as better than others was at odds with the regression effect, and it appears that overestimates of their own performance reduced the WTA effect on the difficult test to nonsignificance.

GENERAL DISCUSSION

The more difficult the task, the more people tend to overestimate their absolute performances (Kruger & Dunning, 1999; Lichtenstein & Fischhoff, 1977). This hard/easy effect can be accounted for by regression effects, simply because any error in estimations of extreme performances will underestimate performances at the top and overestimate performances at the

bottom of the range (Erev et al., 1994; Krueger & Mueller, 2002). Here, we have shown that such regression effects can also account for BTA and WTA effects in comparative judgment. Our results demonstrate that BTA and WTA effects are attributable in part to greater error and greater regressiveness in estimating performance of others than the self. As a result, people estimate the behavior of others as less extreme than their own, resulting in BTA effects on simple tasks and WTA effects on difficult tasks.

The evidence we present here suggests that regression should replace differential weighting as the most viable psychological mechanism to account for BTA and WTA effects. We have presented findings suggesting that differential weighting was not an important factor in the comparative judgments of our participants, and that with clear comparative measures there need not be any difference between direct and indirect measures of comparison. The regression explanation offers a parsimonious theory of the underlying psychological processes that produce these biases in comparative judgment. As we have shown in the second experiment, this process is compatible with focalism accounts. Regression may also be able to account for some of the evidence that has led people to hypothesize that individual-to-group comparison problems are the cause for BTA and WTA effects. To the extent that the referent is vague or people are uncertain about the referent, their estimates of the referent ought naturally to regress to an “ignorance prior.” Indeed, this is the argument made by Giladi and Klar’s (2002) LOGE account: that when people are asked to compare a single target with a group referent, that their conceptualization of the group is less precise and more regressive.

The regression explanation questions the normative status of BTA and WTA effects. On the one hand, participants’ estimates of absolute performance for self and other are clearly imperfect, particularly so for estimates of others. On the other hand, after just having done well,

it is reasonable to assume that others might not do quite as well. After all, any extreme outcome is likely to be unique; other outcomes are likely to be less extreme. Our aim here is to demonstrate a parsimonious explanation that is sufficient to account for BTA and WTA effects. It is not our goal to suggest that people are perfectly rational or that BTA and WTA effects are not real. These effects are real, reliable, and robust. It would be difficult to argue that even perfectly rational people should be able to eliminate any error in their estimates of performance or that they should assume others will be exactly the same as they are. Even so, it is clear that BTA and WTA effects are biases, with potentially profound effects on behavior (e.g., Moore & Cain, 2004).

One of the key assumptions underlying the regression explanation is the “ignorance prior.” The basic assumption is that when they are asked to make some estimate or evaluation of performance, people often arrive with expectations, however vague, about with the distribution of outcomes will look like. Throughout this paper we have assumed that this ignorance prior on an unfamiliar test was a score of 50%. But this is surely not universally true. One obvious exception would be students’ expectations about the distribution of test scores. Many students have gotten used to grading systems in which 90-100% is an A, 80-90% is a B, and so on. This distribution is, of course, highly skewed. Exams that are more difficult—say, with a mean of 50%—are better tests, psychometrically speaking (because ceiling effects are less likely to be a factor). However, any instructor who has used an exam on which students score an average of 50% knows that it results in widespread consternation, even when students know the exam will be graded on a forced curve. Students have difficulty shaking the belief that they will be getting a bad grade and feel sure that they will be below average, even when they have scores above

50%. This is easily explainable by virtue of the fact that their prior belief about the test mean was closer to 80%.

A great deal of research has examined the phenomenon of overconfidence. Many researchers focus on overestimates of one's own actual performance and call it overconfidence (Erev et al., 1994; Griffin, Dunning, & Ross, 1990; Polivy & Herman, 2002; Soll & Klayman, 2004). Others focus on overestimates of one's performance relative to others and call that overconfidence (Bazerman & Neale, 1982; Camerer & Lovallo, 1999; Cooper, Woo, & Dunkelberg, 1988; Neale & Bazerman, 1985). The results presented here highlight the importance of distinguishing between absolute overconfidence and relative overconfidence. The two effects tend to be influenced in opposite ways by task difficulty and the two respond differently to information about performance by self and others. While motivational effects are likely to influence both types of confidence judgments similarly, information influences on confidence judgments differ dramatically for absolute and relative confidence. The regression effect can reconcile the apparent inconsistency between evidence on the hard/easy effect and evidence of BTA and WTA effects.

The regression explanation can help make sense of a set of inconsistent findings in research on risk perception. Research indicates that people tend to overestimate, sometimes radically, the probability that they will experience rare events. For example, Lerner, Gonzalez, Small, and Fischhoff (2003) reported that after September 11th, 2001, Americans estimated their probability of being injured in a terrorist attack as 20%. In a separate domain, women have been found to overestimate the probability that they will fall ill with breast cancer, often by as much as eight times the true probability (Lipkus, Biradavolu, Fenn, Keller, & Rimer, 2001). At the same time, people often believe that their risk for experiencing these rare events is below average

(Woloshin, Schwartz, Black, & Welch, 1999). The regression explanation can account for both these findings.

We have offered an explanation for the psychological mechanisms underlying BTA and WTA effects using a parsimonious explanation that accounts for a significant proportion of the variance in comparative judgment. While our explanation can account for BTA and WTA effects, we do not believe it rules out other effects. For instance, it is clear that some BTA and WTA effects are sometimes due to sloppy measures that lead respondents to conflate relative and absolute evaluation (Moore, 2004). Furthermore, evaluations are partly driven by motivation: People prefer to believe that they are both better than they actually are (Budescu, Wallsten, & Au, 1997) and better than others (Allison et al., 1989; Kunda, 1990). We are not prepared to say that differential weighting of target and referent never occurs in comparative judgment, only that we did not find much evidence for it in our experiments here. Our results do suggest that many previous findings which have been attributed to differential weighting may in fact be entirely attributable to regression and to the use of vague subjective measures. In any case, the regression explanation remains the most parsimonious account of the observed pattern of overconfidence in relative performance and underconfidence in absolute performance on simple tasks and the opposite pattern on difficult tasks.

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.
- Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter than others: The Muhammad Ali effect. *Social Cognition*, *7*(3), 275-295.
- Baron, J. (1997). Confusion of relative and absolute risk in valuation. *Journal of Risk and Uncertainty*, *14*, 301-309.
- Bass, B. M., & Yammarino, F. J. (1991). Congruence of self and others' leadership ratings of naval officers for understanding successful performance. *Applied Psychology: An International Review*, *40*(4), 437-454.
- Bazerman, M. H., & Neale, M. A. (1982). Improving negotiation effectiveness under final offer arbitration: The role of selection and training. *Journal of Applied Psychology*, *67*, 543-548.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, *66*(1), 5-20.
- Budescu, D. V., Wallsten, T. S., & Au, W. T. (1997). On the importance of random error in the study of probability judgment: Part II. Applying the stochastic judgment model to detect systematic trends. *Journal of Behavioral Decision Making*, *10*(3), 173-188.
- Camerer, C. F., & Lovo, 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).
- Chambers, J. R., Windschitl, P. D., & Suls, J. (2003). Egocentrism, event frequency, and comparative optimism: When what happens frequently is "more likely to happen to me". *Personality and Social Psychology Bulletin, 29*(11), 1343-1356.
- Cooper, A. C., Woo, C. Y., & Dunkelberg, W. C. (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing, 3*(2), 97-109.
- Dawes, R. M. (1998). Behavioral decision making and judgment. In D. T. Gilbert & S. T. Fiske (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 497-548). Boston, MA, USA: McGraw-Hill.
- Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes, 65*(3), 201-211.
- 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.
- Eiser, J. R., Pahl, S., & Prins, Y. R. A. (2001). Optimism, pessimism, and the direction of self-other comparisons. *Journal of Experimental Social Psychology, 37*(1), 77-84.
- El-Gamal, M. A., & Grether, D. M. (1995). Are people Bayesian? Uncovering behavioral strategies. *Journal of the American Statistical Association, 90*(432), 1137-1145.

- Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology, 79*(6), 861-875.
- 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.
- Friedman, D. (1998). Monty Hall's three doors: Construction and deconstruction of a choice anomaly. *American Economic Review, September*, 933-946.
- Gigerenzer, G., Hoffrage, U., & Kleinboelting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*(4), 506-528.
- Giladi, E. E., & Klar, Y. (2002). When standards are wide of the mark: Nonselective superiority and inferiority biases in comparative judgments of objects and concepts. *Journal of Experimental Psychology: General, 131*(4), 538-551.
- Griffin, D. W., Dunning, D., & Ross, L. (1990). The role of construal processes in overconfident predictions about the self and others. *Journal of Personality and Social Psychology, 59*(6), 1128-1139.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430-454.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

- 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.
- Klein, W. M. P. (2001). Post hoc construction of self-performance and other performance in self-serving social comparison. *Personality and Social Psychology Bulletin, 27*(6), 744-754.
- Krueger, J. (2000). The projective perception of the social world: A building block of social comparison processes. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 323-351). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Krueger, J. (2003). Return of the ego--Self-referent information as a filter for social prediction: Comment on Karniol (2003). *Psychological Review, 110*(3), 585-590.
- Krueger, J., Acevedo, M., & Robbins, J. M. (in press). Self as sample. In K. Fiedler & P. Juslin (Eds.), *In the beginning there is a sample: Information sampling as a key to understand adaptive cognition*. New York: Cambridge University Press.
- Krueger, J., & Clement, R. W. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality & Social Psychology, 67*(4), 596-610.

- Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality & Social Psychology, 82*(2), 180-188.
- Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*(2), 221-232.
- Kruger, J., & Burrus, J. (2004). Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology, 40*(3), 332-340.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.
- Kuiper, N. A., & Rogers, T. B. (1979). Encoding of personal information: Self-other differences. *Journal of Personality and Social Psychology, 37*(4), 499-514.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480-498.
- Lerner, J. S., Gonzalez, R. M., Small, D. A., & Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychological Science, 14*(2), 144-150.
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Decision Processes, 20*(2), 159-183.
- Lipkus, I. M., Biradavolu, M., Fenn, K., Keller, P., & Rimer, B. K. (2001). Informing women about their breast cancer risks: Truth and consequences. *Health Communication, 13*(2), 205-226.

- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*(3), 280-296.
- MacDonald, T. K., & Ross, M. (1999). Assessing the accuracy of predictions about dating relationships: How and why do lovers' predictions differ from those made by observers? *Personality and Social Psychology Bulletin, 25*(11), 1417-1429.
- McCall, M., & Matrass, K. (2001). Carding for the purchase of alcohol: I'm tougher than the other clerks are! *Journal of Applied Social Psychology, 31*(10), 2184-2194.
- McFarland, C., & Miller, D. T. (1990). Judgments of self-other similarity: Just like other people, only more so. *Personality and Social Psychology Bulletin, 16*(3), 475-484.
- 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. (2004). When good = better than average. Pittsburgh: Carnegie Mellon. Unpublished manuscript.
- Moore, D. A., & Cain, D. M. (2004). *Myopic biases in comparative judgment and entrepreneurial entry*. Pittsburgh, PA. Tepper Working Paper Number 2003-E76.
- 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., & Klein, W. M. P. (2004). *The use and abuse of absolute and comparative information in judgment*. Pittsburgh, PA.
- Moore, D. A., Kurtzberg, T. R., Fox, C. R., & Bazerman, M. H. (1999). Positive illusions and forecasting errors in mutual fund investment decisions. *Organizational Behavior & Human Decision Processes, 79*(2), 95-114.

- Neale, M. A., & Bazerman, M. H. (1985). The effects of framing and negotiator overconfidence on bargaining behaviors and outcomes. *Academy of Management Journal*, 28(1), 34-49.
- Perloff, L. S., & Fetzer, B. K. (1986). Self-other judgments and perceived vulnerability to victimization. *Journal of Personality & Social Psychology*, 50(3), 502-510.
- Polivy, J., & Herman, C. P. (2002). If at first you don't succeed: False hopes of self-change. *American Psychologist*, 57(9), 677-689.
- Risucci, D. A., Tortolani, A. J., & Ward, R. J. (1989). Ratings of surgical residents by self, supervisors and peers. *Surgery, Gynecology, and Obstetrics*, 169, 519-526.
- Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3), 279-301.
- Smith, V. L., & Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry*, 31(2), 245-261.
- Soll, J. B. (1996). Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes*, 65(2), 117-137.
- Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30(2), 299-314.
- Suls, J. M., Lemos, K., & Stewart, H. L. (2002). Self-esteem, construal, and comparisons with the self, friends, and peers. *Journal of Personality & 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.

- Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *101*(4), 547-567.
- Van Boven, L., Dunning, D., & Loewenstein, G. (2000). Egocentric empathy gaps between owners and buyers: Misperceptions of the endowment effect. *Journal of Personality and Social Psychology*, *79*(1), 66-76.
- Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, *29*(9), 1159-1168.
- 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.
- Woloshin, S., Schwartz, L. M., Black, W. C., & Welch, H. G. (1999). Women's perceptions of breast cancer risk: How you ask matters. *Medical Decision Making*, *19*(3), 221-229.

Footnotes

¹ In order to rule out timing effects of this procedure, half of these participants drew their random number before they bet, half after they bet. Because this manipulation did not yield any main or interaction effects with variables of interest, these participants are grouped together in the analyses reported.

² Degrees of freedom fluctuate between tests due to missing data for some participants.

³ There is more noise and variability in the middle of the scale on Figure 2 because there are so few participants with scores of 5 and 6. Most of the scores lie at the extremes.

⁴ Similar patterns hold for bets and probability of winning as measures of comparative judgment. Bets, however, are influenced by a number of other motivations, including risk aversion and feelings about gambling.

⁵ Those who were focusing on themselves all wrote these essays. They were given the assignment to: “write a page or less about yourself on the topic of ‘what makes me unique.’ As a starting point, please begin the essay by talking about who you are: your gender, your age, your race, your family background, your home, and your major field of study at the university. You may also want to mention your interests, your skills, and your abilities.” The essays that they had written were then read by participants in other conditions.

⁶ These 330 other participants were part of another study, the results of which are reported elsewhere (Moore & Klein, 2004).

⁷ Note that these same effects are replicated in other measures of comparative judgment: bets and rated relative performance.

⁸ It is noteworthy that in the no-feedback condition participants in the difficult condition did not think they were below average. On average, they believed that their scores would be .31 above average ($SD = 2.99$), which is not significantly different from zero, $t(35) = .61, ns$. The reduction of the WTA effect in the difficult condition and the BTA effect in the simple condition is consistent with the regression explanation. When they completely lack information about themselves, people's performance estimates for themselves are likely to be regressive, as they are for others.

⁹ Significance tests do not apply to these results because the independent variables perfectly account for all the variation in the dependent variable.

¹⁰ The main between-subjects effect of self-feedback is also significant, $F(1, 331) = 10.18, p = .002, \eta^2 = .03$. It describes the tendency for self-feedback to moderate people's estimates of their own performance. This finding is discussed more later under the heading of "self-enhancement."

¹¹ Naturally, the main between-subjects effect of difficulty is also significant, $F(1, 333) = 514.09, p < .001$.

Table 1

Predicted and actual scores by experimental condition in Experiment 1. Standard deviations in parentheses. Figures on the same row with different superscripts are significantly different from each other ($p < .05$).

	<u>Betting on self</u>		<u>Betting on a random person</u>	
	Simple	Difficult	Simple	Difficult
Actual score (out of 10)	8.58 ^a (1.59)	1.76 ^b (1.34)	8.48 ^a (1.43)	1.60 ^b (1.22)
Bet (up to \$3)	\$2.13 ^a (\$.88)	\$1.41 ^b (\$1.23)	\$1.69 ^{a,b} (\$1.05)	\$1.37 ^b (\$1.15)
Probability of winning	64% ^a (20%)	37% ^c (24%)	51% ^b (21%)	39% ^c (21%)
Comparative evaluation (7-point scale)	5.05 ^a (.99)	3.34 ^d (1.24)	4.25 ^b (.84)	3.85 ^c (.69)
Protagonist's score (estimated)	8.37 ^a (1.46)	2.22 ^c (1.34)	7.66 ^a (1.48)	3.32 ^b (1.62)
Opponent's score (estimated)	7.68 ^a (1.02)	3.06 ^b (1.41)	7.32 ^a (1.64)	3.37 ^b (1.57)
Average score (estimated)	7.12 ^a (1.11)	3.07 ^b (1.16)	7.05 ^a (1.48)	3.29 ^b (1.33)
Size of 90% conf. interval—protagonist	2.79 ^a (1.43)	3.09 ^a (1.31)	4.33 ^b (2.05)	4.69 ^b (1.94)
Size of 90% conf. interval—opponent	3.92 ^a (1.75)	4.39 ^a (2.15)	4.13 ^a (2.09)	4.89 ^a (2.14)

Table 2

Results of regressions on participants' bets using predictions of absolute performance of protagonist and opponent as independent variables, comparing those focused on protagonist and those focused on opponent (from Moore & Kim, 2003, Experiment 4).

	<u>Bets</u>	
	Focusing on Protagonist	Focusing on Opponent
Protagonist's score (0-10)	1.51***	.53†
Opponent's score (0-10)	-.76**	-.97**
Adjusted R-square	.70***	.26***

Note: Table shows standardized beta-weights for independent variables.

†p<.10

**p<.01

***p<.001

Table 3

Results for the five different measures of comparative judgment, Experiment 3. The second column shows the effect size of the difference between simple and difficult conditions, as well as the significance of the t-test comparing the two conditions. Regression results predicting comparative judgment based on absolute evaluations of self appear in the third and fourth columns.

Dependent Measure of Comparative Judgment	Simple vs. Difficult	Regression results	
	Effect size (η^2)	β (Self)	β (Other)
Estimated difference between self and others	.05***	1.26***	-.96***
Indirect comparison (implicit)	.07***	1.67 †	-1.38 †
Bet on self being above average	.13***	.95***	-.55***
Relative rating (1-7 scale)	.17***	.99***	-.52***

*** $p < .001$

† DV is perfectly determined by IVs.

Table 4

A comparison of influences on comparative judgments in four experimental conditions, Experiment 3. The second column shows the effect size of the difference between simple and difficult conditions, as well as the significance of the t-test comparing the two conditions. Regression results predicting comparative judgments, based on absolute evaluations of self and other appear in the third and fourth columns.

Condition	Simple vs. Difficult Effect size (η^2)	Regression results	
		β (Self)	β (Other)
No feedback about self or others	.02	.93***	-.69***
Other feedback only	.01	1.00***	-.99***
Self feedback only	.31***	1.22***	-.66***
Full feedback about self and others	.06*	1.67***	-1.39***

* $p < .05$

*** $p < .001$

Figure 1. Estimated scores for self and other (hypothetical data).

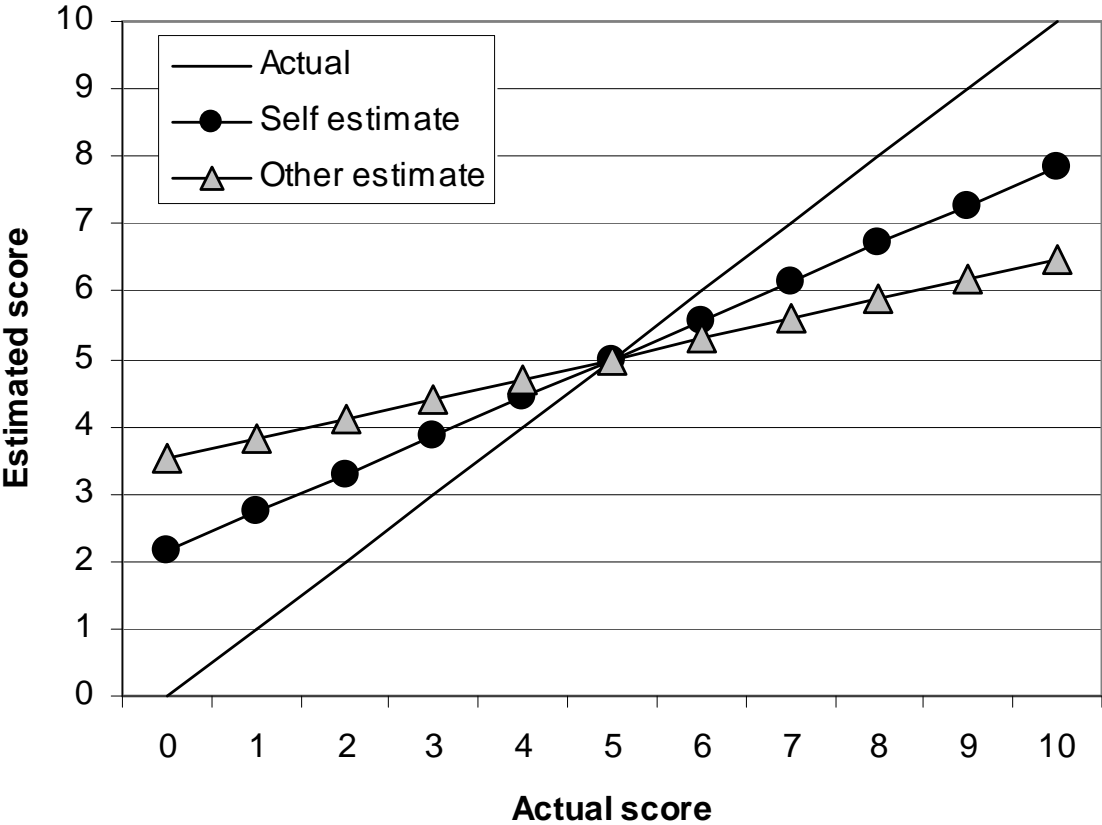


Figure 2. Estimated scores for self and opponent by those betting on the self, Experiment 1

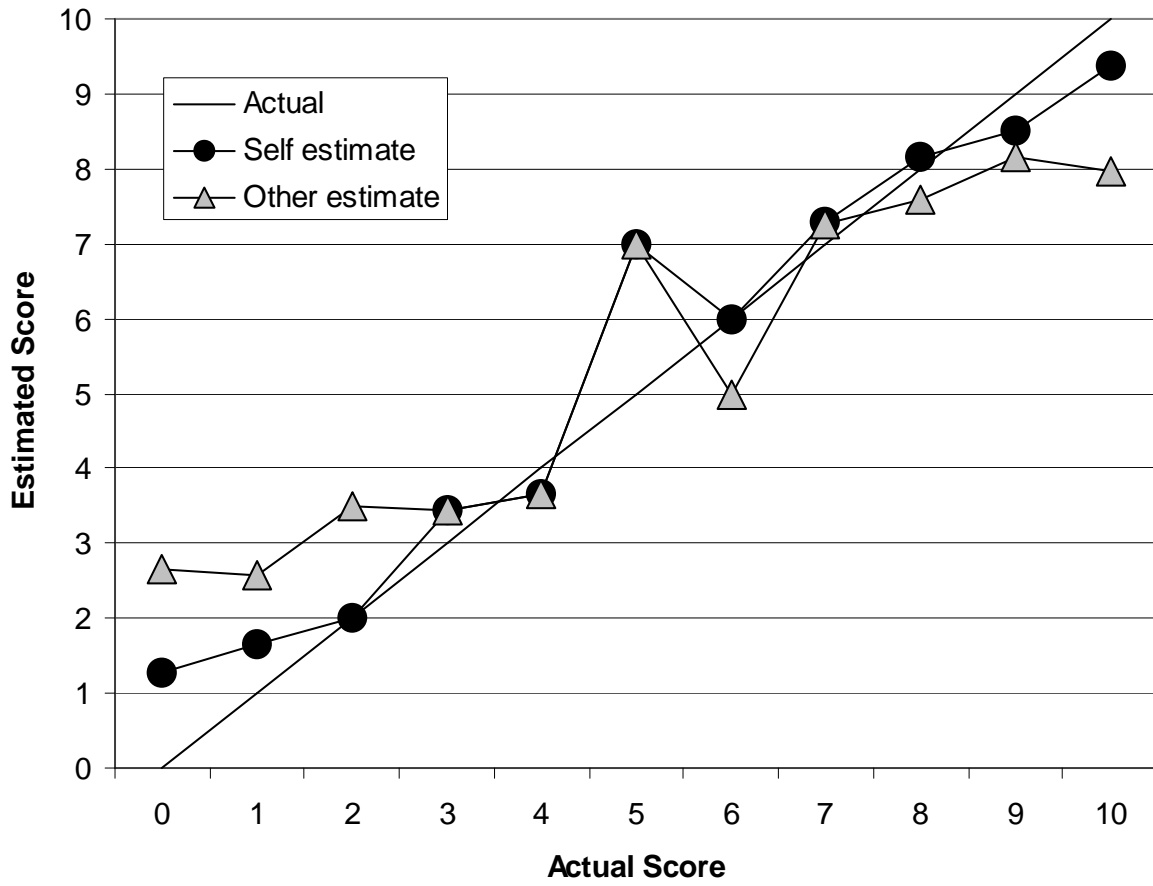
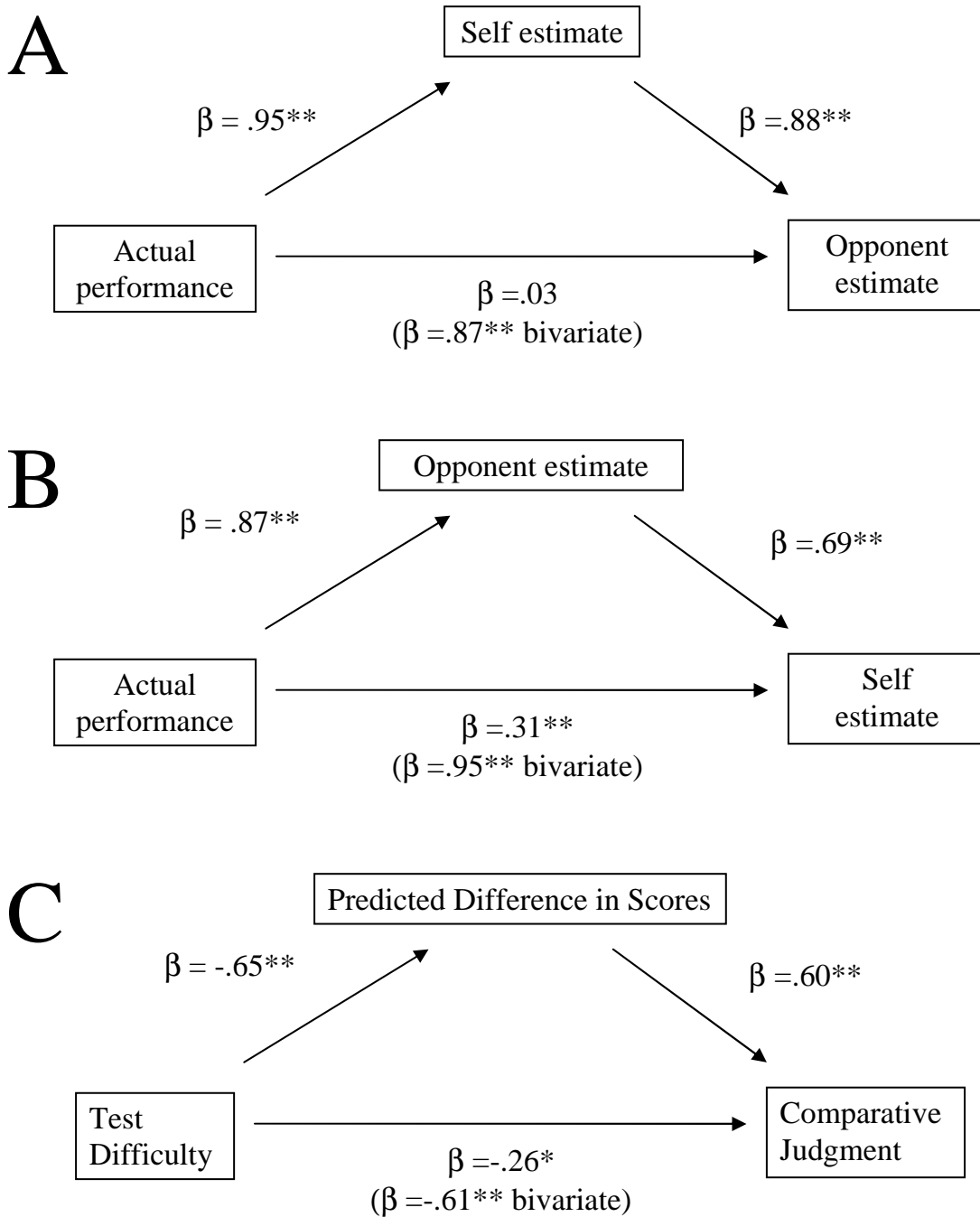


Figure 3. Mediation analyses (among those betting on themselves), Experiment 1.



* $p < .05$, ** $p < .001$

Figure 4. Mediation analyses examining how participants' actual performances influenced their estimates of performance by themselves and others, as a function of whom they were focusing on, Experiment 2.

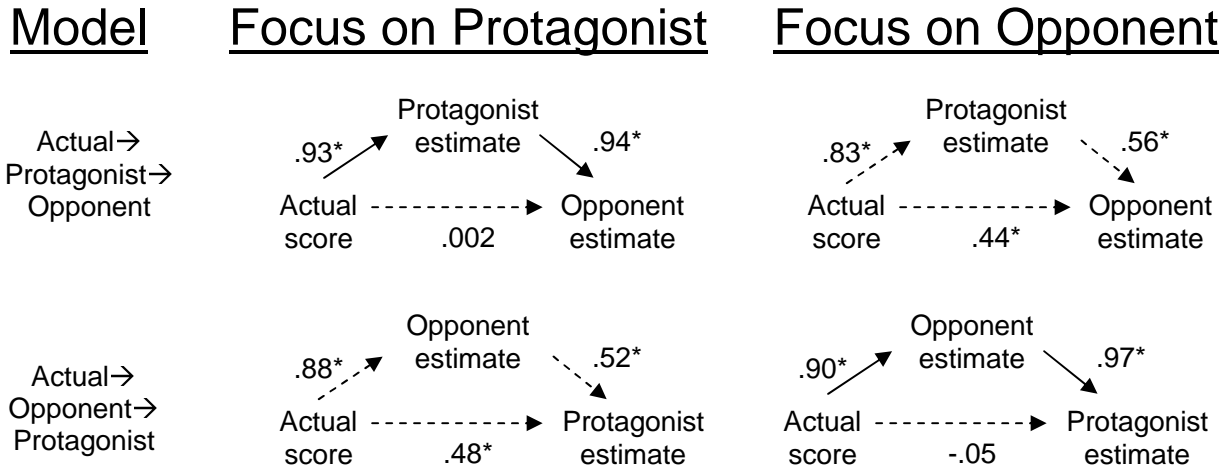


Figure 5. Participants' estimates of their performance relative to the average, as a function of task difficulty and feedback about their own performance in Experiment 3. Estimates are relative to the average, so a rating of 1 indicates that participants in that condition believed they scored 1 point above average. Negative scores indicate that people thought they were below average. Error bars show standard errors.

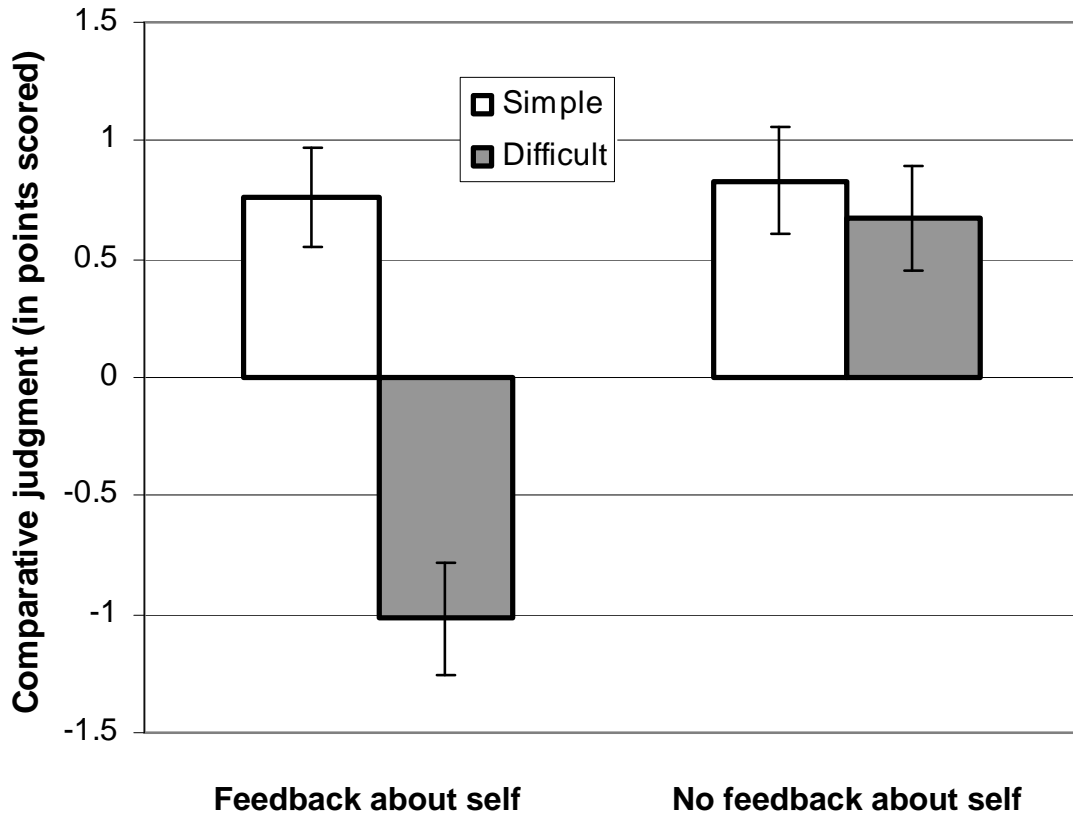


Figure 6. Participants' estimates of their performance relative to the average, as a function of task difficulty and feedback about their own performance in Experiment 3. Estimates are relative to the average, so a rating of 1 indicates that participants in that condition believed they scored 1 point above average. Negative scores indicate that people thought they were below average. Error bars show standard errors.

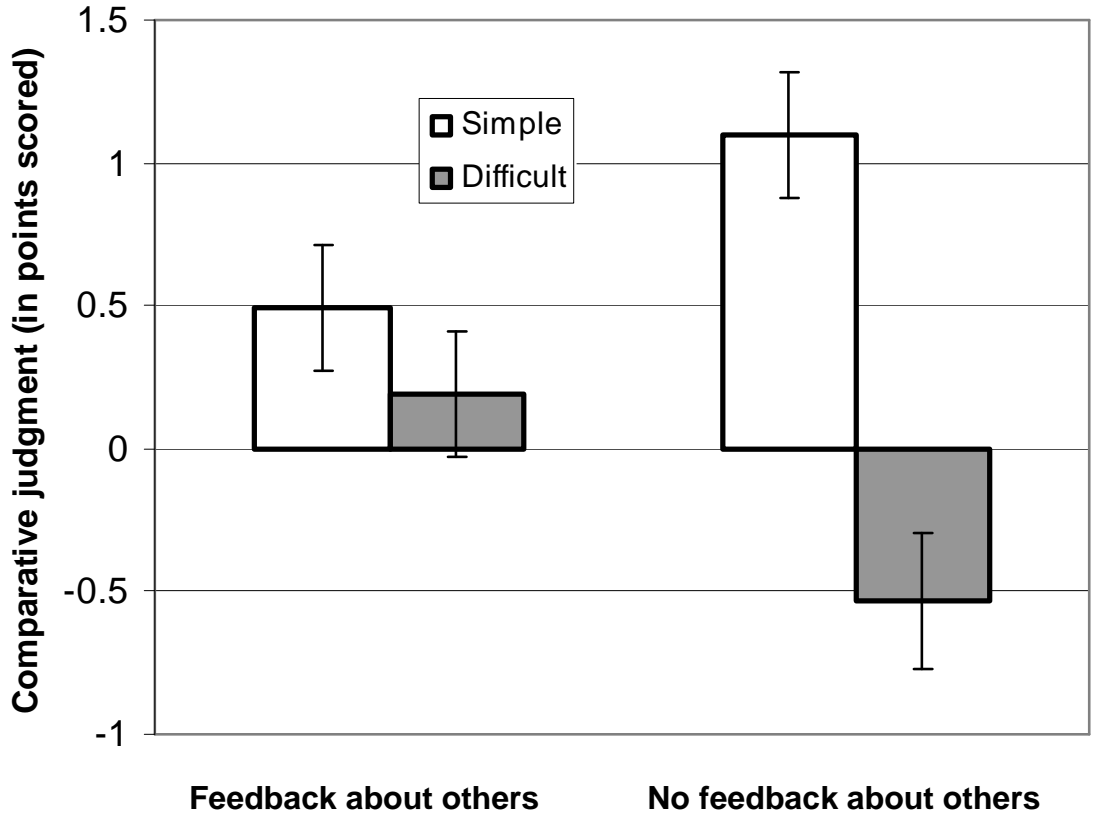
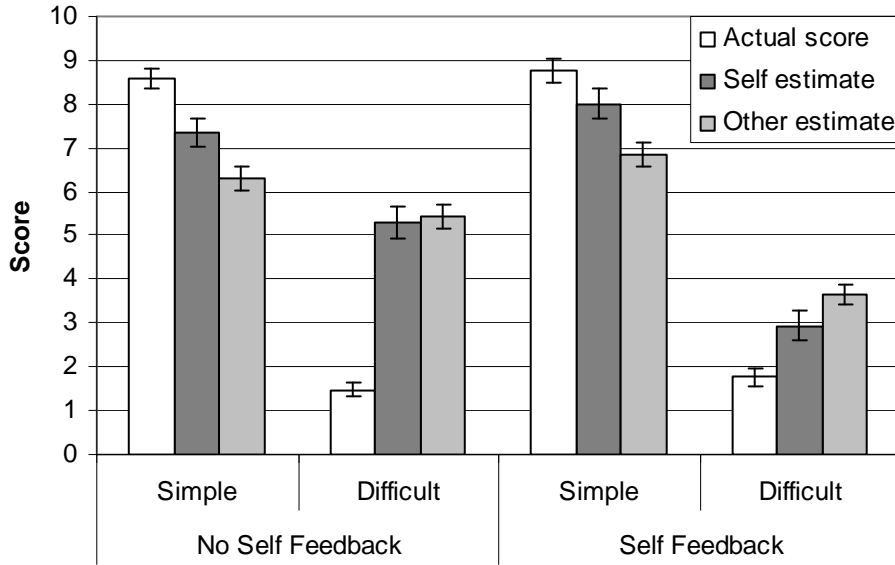


Figure 7. Participants' estimates of absolute performance of self and other, as a function of task difficulty and feedback about their own performance in Experiment 3. Error bars show standard errors.

Panel A: No Feedback About Others



Panel B: Feedback About Average Score

