Running head: COMPARATIVE JUDGMENT

NOTE: This is not the copy of record. This paper has appeared in print:
Moore, D. A., & Small, D. A. (2007). Error and bias in comparative social judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology, 92*(6), 972-989.

Error and Bias in Comparative Judgment:

On Being Both Better and Worse Than We Think We Are

Don A. Moore
Carnegie Mellon University

Deborah A. Small
University of Pennsylvania

Abstract
People believe that they are better than others on easy tasks and worse than others on difficult tasks. Previous attempts to explain these better-than-average and worse-than-average effects have invoked bias and motivation as causes. This paper develops a more parsimonious account, the differential information explanation, which assumes only 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 easy tasks and below average on difficult tasks. These effects are exacerbated when people have accurate information about their performances, increasing the natural discrepancy between knowledge of self and others. The effects are attenuated when people obtain accurate information about the performances of others.

The paper benefited from constructive comments by Monica Biernat, Peter Boatwright, John Chambers, Jason Dana, Robyn Dawes, Klaus Fiedler, Joachim Krueger, Thomas Mussweiler, Bryan Routledge, Zach Sharek, Bill von Hippel, and Paul Windschitl. Thanks also to Zach Burns, Joshua Furgeson, Sam Swift, Sapna Shah, and Wemi Peters for help with data collection. The authors appreciate the support of National Science Foundation Grant SES-0451736, a Berkman Junior Faculty Development grant from Carnegie Mellon University, and the staff and facilities of the Center for Behavioral Decision Research. Please send correspondence to don.moore@alumni.carleton.edu.

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

How good are we at assessing ourselves? 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 (Dunning, 2005; Dunning, Heath, & Suls, 2004; Epley & Dunning, 2000; Messick, Bloom, Boldizar, & Samuelson, 1985; Moore, Kurtzberg, Fox, & Bazerman, 1999; Svenson, 1981; Weinstein, 1980). These better-than-average (BTA) beliefs have profound implications because they are fundamental to competitive decision making in all kinds of contexts. If entrepreneurs believe they are better than others, that would contribute to the high rates of entrepreneurial entry (Camerer & Lovallo, 1999; Cooper, Woo, & Dunkelberg, 1988). If CEOs believe that they are better than other CEOs, it would contribute to higher rates of corporate acquisition in which firms buy others that they believe they can manage more effectively (Malmendier & Tate, 2004, 2005). And if disputants believe that they are better than their opponents, it could help explain the frequency of strikes, lawsuits, and wars (Babcock & Olson, 1992; Howard, 1983; Neale & Bazerman, 1985).

Yet recent evidence suggests that people do not always believe they are better than others. Indeed, people consistently rate themselves *below* average in some domains (Kruger, 1999; Moore & Kim, 2003; Windschitl, Kruger, & Simms, 2003). People report themselves to be below average in juggling, their probability of living past 100, and their ability to cope with the death of a loved one (Blanton, Axsom, McClive, & Price, 2001; Chambers, Windschitl, &

Suls, 2003; Kruger, 1999; Kruger & Burrus, 2004). The characteristic feature of such worse-than-average (WTA) effects is that they occur in domains where success is rare. Most people do not live past 100 or carry on happily after the death of a loved one. By contrast, domains in which people tend to rate themselves better than average are domains in which people generally feel capable. In this paper, we explore the possibility that these effects are in part due to the simple fact that people possess better information about themselves than others. Our theory is that, given that people have more information about themselves than about others, when their own performances are exceptional (either good or bad), it is reasonable for them to assume others' will be less exceptional. Consequently, they will estimate the performances of others as less extreme than their own. The result is that they will believe they are above average on tasks where they have performed well and below average tasks where they have performed poorly.

Prior Explanations for BTA and WTA effects

Motivated reasoning

BTA effects have frequently been explained using the fact that people are motivated to view themselves in a positive light. For instance, Taylor and Brown (1988) argued that people prefer to believe in their own superiority because these positive illusions help them persist in the face of life's many frustrations, and that they even promote mental health. However, motivational accounts generally have trouble explaining WTA effects. Taylor and Brown's theory of positive illusions does not offer a compelling explanation for why people believe they are less likely than others to live past 100.

Differential weighting

The leading non-motivational explanations all suggest that comparative judgments focus disproportionately on the target (usually the self) at the expense of consideration of the referent to whom the target is being compared (for reviews see Chambers & Windschitl, 2004; Moore, in press). For example, Klar and Giladi (1999) argued that when people are asked to compare themselves with others, their reports are, in effect, just self-evaluations. They showed that the degree to which people report themselves as happier than others is highly correlated with their own self-reported level of happiness, but is only weakly correlated with their estimates of others' happiness. The basic idea is that people overweight their own happiness and underweight others' happiness when judging whether they are happier than others.

Researchers have demonstrated this differential weighting effect using path analyses. These analyses are conducted as follows: Individual estimates of self and others serve as independent variables, and comparative judgments of the self relative to others serve as the dependent variable. The standard result is that self-assessments are strongly correlated with comparative judgments but that assessments of others are not (see Klar & Giladi, 1997). This result has been interpreted to mean that individual assessments of the self are overweighted in comparative judgment. There are two problems with this interpretation.

Problems with the Differential Weighting Explanation

First, comparative judgments *ought* to share more variance with one's own than with others' performance. Usually, researchers use a group (such as other students at the same university) as the referent other. If all respondents correctly estimated the group mean, then their reports would show no variation and would thus be uncorrelated with their comparative judgments. Imagine I ask the members of one of my classes to report whether they are taller or shorter than average for the class. If everyone in class accurately estimates both their own height and the class's average height, then the standard path analysis would find (1) that class members' reports of their own heights would account for 100% of the variance in their comparative judgments and (2) that their estimates of the class's average height, which includes no variance, would account for none (0%) of the variance in their comparative judgments. In sum, if everyone is responding accurately, then the path analysis would make it appear as if the target was weighted more heavily than referent.

The second problem with the standard path analysis is that comparative judgments are often elicited using vague subjective measures that are easily conflated with individual evaluations of the self (Biernat, Manis, & Kobrynowicz, 1997; Burson & Klayman, 2005; Moore, 2006). This conflation inflates the covariation between individual and relative evaluations of the self, making individual self-assessments appear more predictive of comparative self-assessments.

A distinct weakness of the differential weighting explanation is that it offers no account for how people form individual assessments of target and referent. The differential weighting can explain the misuse of individual assessments once they are made. However, it does not hypothesize a systematic difference between individual assessments of self and other, and it does not provide an explanation regarding how people make individual assessments. Differential weighting, therefore, offers no explanation for BTA and WTA effects implicit in individual assessments (so-called "indirect" comparative judgments, computed by subtracting individual estimates of the self from individual estimates of a referent). Differential weighting only offers to explain how these individual assessments might be used (or misused) to form direct comparative judgments, and only predicts BTA and WTA effects on direct comparative judgments.

The Differential Information Explanation for BTA and WTA effects

This paper examines another potential cause of BTA and WTA effects: differential regression. In particular, we focus on differential information about target and referent as a cause for greater regressiveness in estimates of others. This theory can account for the data on BTA and WTA effects, and it can also explain the origins of individual assessments of self and others. Importantly, it can account for the presence of BTA and WTA effects in individual assessments, which differential weighting theories cannot.

Our explanation begins with the fact that people usually possess better information about themselves than about others (Dawes & Mulford, 1996; Epley & Dunning, 2006; Fiedler, 1996; Karniol, 2003; Keysar, Barr, Balin, & Brauner, 2000; Krueger, 2000; Mussweiler, 2003; Ross & Sicoly, 1979; Van Boven, Dunning, & Loewenstein, 2000; but for some noteworthy exceptions, see Bass & Yammarino, 1991; MacDonald & Ross, 1999; Risucci, Tortolani, & Ward, 1989). This simple fact has profound implications. One of them is that people's estimations of others are less extreme than their estimations of themselves (Miller & McFarland, 1987; Sande, Goethals, & Radloff, 1988). This implication must follow, statistically. Individual outcomes will vary around a group average. Without individuating information about another person's outcomes, if one were asked to estimate that other individual's outcome, the group's average would be a good opening assumption (or what statisticians call a *prior*). However, people who know their own outcomes also know the degree to which they deviate from the group's average.

Greater information about the self means that people can make more informed estimates of their own outcomes in life, such as how much they weigh, their honesty, or their risk of committing suicide. Estimates of others' outcomes, however, since they are based on less information, must rely more heavily on guesses about group base rates and average outcomes.

Our theory hypothesizes that BTA and WTA effects result from the inability to sufficiently update priors due to inadequate information about others.

For instance, when people are asked to estimate whether they tell the truth more often than others, they must guess the frequency with which others lie. Most people tell the truth most of the time. However, none of us can be equally sure that others try as hard as we do to be honest. For frequent events like truth-telling, it is far easier to underestimate the actual base rate than it is to overestimate it. The result is that most of us believe that we are more honest than others. On the other hand, suicidal thoughts are rare. Since people are aware of their own (lack of) suicidal thoughts but cannot be so sure about others, it is common for people to believe that their own likelihood of committing suicide is lower than is that of others.

How Exactly Does the Differential Information Theory Explain BTA and WTA Effects?

To clarify our theory, let us consider an easily quantifiable example. Imagine that you have taken a test on which you knew all of the correct answers. You know you did well, but you do not know how others did. How did you do relative to others? Unless you believe 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 everything right is less than 100%. As such, your estimate of others' 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.

This hypothetical pattern of results is illustrated in Figure 1. Figure 1 illustrates two key empirical regularities: (1) that people have imperfect knowledge of their own performances, and so make regressive estimates of themselves (Burson, Larrick, & Klayman, 2006; Erev, Wallsten, & Budescu, 1994); and (2) that their estimates of others are even more regressive. In this way, our theory explains the inconvenient co-occurrence of two apparently contradictory findings: The first is that people think that they are better than others on easy tasks and worse than others on difficult tasks. The second is that people are most likely to overestimate their actual performances when the task is difficult (Krueger & Mueller, 2002; Lichtenstein & Fischhoff, 1977; Lichtenstein, Fischhoff, & Phillips, 1982). We document both these patterns in all three of our experiments.

Note that our theory would predict a reversal of the standard BTA/WTA effects when people have better information about others than about the self. To be more precise, the provision of information about others' outcomes will make people more likely to believe that they are worse than others on easy tasks and more likely to believe that they are better than others on difficult tasks. This is a novel prediction that distinguishes our theory from others. We test this prediction in our second and third experiments.

Theoretical Foundations of our Differential Information Theory

Fiedler (1996; 2000) suggested better information about the self than others as an explanation for BTA effects. The logic of his argument is quite consistent with our own. Fiedler's model begins with the observation that people usually have more information about themselves than they do about others, and observe more instances of their own behavior than that of others. He points out that, since desirable behaviors such as being cooperative and friendly are generally more frequent than undesirable behaviors such as being rude and phony, people will observe themselves engaging in more desirable behaviors than others. As such, it often makes sense for them to conclude that positive traits are more descriptive of themselves than of others (e.g., Alicke, 1985). Like Fiedler's theory, ours does not confer any special role to the self, other than the fact that people happen to have more information about themselves than about others. We build on Fiedler's work in three ways. First, we use it to help explain recent evidence of *worse*-than-average effects. For domains in which failure is more frequent than success, people will more frequently observe their own failures than others' failures and infer that they are worse than others.

Second, our theory elaborates on Fielder's model by explicitly allowing for the possibility that people use information about themselves to make inferences about others. The tendency to assume that others are like us has been called the "false consensus effect" (L. Ross, Greene, & House, 1977). However, Dawes (1989) has pointed out that there is a defensible normative basis for using information about ourselves to make inferences about others. Let's say, for instance, that an unfortunate camper learns the hard way that poison ivy does not make good toilet paper. What will she tell her fellow campers? "Try it out for yourself. It wasn't great for me, but who knows—it might be just your cup of tea"? More likely, she will make the sensible inference that wiping themselves with poison ivy would be a mistake for others as well. People do in fact use themselves as a basis for helping them understand others (Krueger & Clement, 1994) and they should.

However, people certainly don't believe that others are exactly like them. If they did, there would be no BTA or WTA effects. But the empirical evidence clearly shows that the self is used as a handy, albeit imperfect, predictor of others (Krueger, 2000; Krueger, Acevedo, & Robbins, 2005; McFarland & Miller, 1990; Mussweiler, 2003). People believe, for instance, that annoying music annoys them more than it does others, and that good music gives them more joy than it does others (Chambers & Suls, in press), or that poison ivy gives them more pain than it would others. At this point, the third contribution of our theory becomes critical: the prior.

People update from their prior beliefs more for themselves than for others, and of course this makes sense when they have more information useful for updating beliefs about the self than about others. [1] This differential updating from prior beliefs can produce the co-occurrence of false consensus (in which people believe that others are more like them than they actually are) and false uniqueness (in which people believe they are more unique than they actually are). For example, Moore and Kim (2003) demonstrated that their participants' beliefs about their own scores on a trivia test were highly predictive of their beliefs regarding others' scores, consistent with false consensus. At the same time, people believed that their own scores would be more exceptional than they actually were: Those who took the easy test believed that they scored better than others, while those who took the difficult test believed they scored worse than others. This pattern is illustrated in Figure 1. Our contention is that beliefs about performance are frequently imperfect, and so often regress from actual performance toward people's prior expectations for performance. In the experiments we present, we test this contention, as well as the hypothesis that more accurate information reduces this regressiveness.

Some Supportive Evidence

Information about others is more observable on some tasks than on others. We have more information, for instance, about others' height (which we can observe directly) than their honesty (which we cannot). Our theory would predict that biases in comparative judgment would be strongest when others' outcomes are difficult to observe. Prior research is consistent with this hypothesis (Allison, Messick, & Goethals, 1989; Miller & McFarland, 1987). Paunonen (1989), for example, showed that BTA effects for common, desirable traits were stronger for unobservable traits than for observable traits, and that this effect was stronger when people were comparing themselves with strangers than with close friends.

Other Contributors to Differential Regressiveness

Evidence suggests that differential regressiveness also depends on referent salience—people display stronger BTA and WTA effects when comparing themselves with vague or abstract others (see Giladi & Klar, 2002; Klar, 2002; Klar & Giladi, 1997). These results have previously been attributed to differential weighting, but differential weighting cannot explain evidence that people make more regressive *individual* estimates of vague or abstract others (Chapman, 1967; Fiedler, 1991, 2000; Sanbonmatsu, Shavitt, & Gibson, 1994; Sanbonmatsu, Shavitt, Sherman, & Roskos-Ewoldsen, 1987). Our theory can account for these results if people are more likely to infer that information about their own performances is more informative for the performance of a similar specific other than it is for the performance of a vague or abstract other—and that they therefore have better information with which to update their priors for the specific than for a vague other. Experiments 1 and 3 afford separate tests of the degree to which salience affects both weighting and regressiveness of individual assessments.

Research Overview

In this paper, we present three experiments that test the main predictions made by our differential information theory. Each of the studies also addresses different findings from the research on biases in comparative judgments, and demonstrates how these findings can be parsimoniously reconciled with or explained by the differential information explanation. Each of the studies also allows us to compare specific predictions made by our theory with the predictions made by other competing theories. In the first experiment, we measure beliefs about performance by self and by others, both before and after taking a test. This allows us to measure how information about performance leads to changes in beliefs, and the degree to which people will project beliefs about their own performance on to others. Experiment 1 tests the prediction that BTA and WTA effects are the result of incompletely revised priors, and that people's beliefs about performance regress toward these priors, more so for others than for self. Experiment 1 also examines the effect of focusing by manipulating the salience of the referent.

The second experiment directly manipulates participants' knowledge about their own and others' performances. We show that, consistent with our theory, providing better information about *own* performance exacerbates BTA and WTA effects, whereas better information about *others'* performance reduces BTA and WTA effects. Experiment 3 capitalizes on the fact that our theory does not distinguish between self-other and other-other comparisons. In our theory, the distinguishing feature of the self is that people have more information about it. Experiment 3, therefore, manipulates the salience of the two other people as well as the information that participants have about them. The results replicate the key findings of the first two experiments and suggest that differential information is a stronger influence on BTA and WTA effects than is focusing.

Our results do not show that differential information can account for all BTA and WTA effects. What they do show, however, is that our theory can account for more of the BTA and WTA effects we observe than can any other theory. But more importantly, this explanatory power comes from a normative theory which explains how these effects can arise as a natural consequence of sensible people making rational inferences using imperfect information.

EXPERIMENT 1: THE TRIVIA QUIZ

Experiment 1 tests some of the basic contentions of the differential information account. We elicit participants' expectations for performance by self and others on a 10-item trivia quiz (their priors). They then take a quiz that is either easy or difficult, and we measure their updated beliefs about performance. Naturally, we expect to replicate BTA effects among those taking the easy test and WTA effects among those taking the difficult test. Unlike the differential weighting explanation, our theory holds that these effects will result from people's underlying beliefs regarding individual performance by self and others. Furthermore, unlike other theories of comparative judgment, our theory specifies how it is that people arrive at these individual assessments. We expect that their beliefs about their own performance will be a combination of, and will lie between, their actual performances and their priors. Furthermore, because they have less useful information about others, our theory predicts that estimates of others' performances will regress toward priors more so than will individuals' own estimates of their individual performances.

We show that, consistent with our theory and with normative principles, participants use information about their own performances to update beliefs about performance more for themselves than for others, and that this tendency can explain the effect of test difficulty on participants' beliefs about how they compare with others. We also vary the salience of the others with whom participants compare themselves, as a way of examining focusing as a cause of more regressive estimates of others.

Method

Participants. The 255 participants were primarily undergraduate students from two large private U.S. universities. Females constituted 56% of participants.

Procedure. Participants were seated at computers and instructed to log in to a page on the world wide web that led them through the entire experiment. The first page read: "*Thanks for participating in this experiment! You have earned $4 for your participation. Shortly, you* *will take the trivia test. Your goal will be to get as many questions right as possible in the allotted time. The test will consist of 10 questions. We will ask you to guess your own score, both before and after you take the test*."

Half the participants, those in the high salience condition then read, "*We will also ask you to guess the score of another person who is taking the same quiz at the same time you are. The other person is sitting in the seat next to yours. Please turn to that person now, shake hands, and introduce yourself*."

The other half of participants, those in the low salience condition read, "*We will also ask you to guess the score of the average person. To be more specific, we will ask you to guess the most common single score obtained by the hundreds of people who have taken this trivia quiz. We will refer to this score as the typical score*."

Participants were then asked to estimate the probability that they would obtain each of the eleven possible scores on the trivia quiz. That is, they made eleven probability estimates, one for each score (0 through 10) predicting how likely it is that they personally would get that many questions right on the trivia quiz they were about to take. The instructions conceded, "*We realize that you have very little information useful for answering these first questions, but please answer as best you can*." Participants were then asked to make the same score prediction for the other person (either the person sitting next to them or the typical person). For the 5% of participants whose probability estimates did not sum to 100%, we divided each of the eleven estimates by $s/100$, where s was the sum of all eleven estimates, thereby forcing the summed probabilities to equal 100%.

Participants then took the 10-item trivia quiz, which was either easy (e.g., "*Berlin is the capital of what country?*") or difficult (e.g., "*What is the capital of Azerbaijan?*"). The 10 trivia items are listed in the Appendix. Participants were then again asked to estimate the probability that they and the other had obtained each of the 11 possible scores. Measuring participants' subjective probability distributions in this way is not traditional, but it is useful for assessing how participants' update their beliefs over the range of possible outcomes.

Then participants were asked the following series of questions, presented to each participant in a different randomly-determined order. Note that, as described below, individual and comparative evaluations were elicited using subjective verbally anchored scales, as was been standard practice, as well as using more objective measures. Inclusion of the traditional measures is useful for reconciling our results with the results of prior research.

Every possible ordering of the following questions was equally likely:

1. "*How many points above or below the other/typical person's do you think your score will be?"* We will refer to this as the direct comparative measure (c.f. Chambers & Windschitl, 2004).
2. "*How do you expect that you will score relative to all the other people taking the same test?*" Participants were provided with a seven-point scale with labels at 1 (*well below average*), 4 (*same*), and 7 (*well above average*).
3. "*How likely do you think it is that your score will be higher than that of the other/typical person? (Between 0% and 100%)*"
4. "*What percentage of the participants in this experiment will have scores below yours? (If you expect your score will be the very best, then put 100. If you expect your score will be exactly in the middle, put 50. If you expect your score will be the lowest, put 0.)*"
5. "*How many of the 10 items do you think you got right?*"

6. "*How many of the 10 questions do you predict that the other/typical person will answer correctly?*"
7. "*How well do you think you did on the quiz?*" Participants were provided with a seven-point scale with labels at 1 (*very poorly*) and 7 (*very well*).
8. "*How well do you think the other/typical person did on the quiz?*" Participants were provided with a seven-point scale with labels at 1 (*very poorly*) and 7 (*very well*).
9. "*How do you choose to bet?*" Participants had to allocate all of their $4 earnings between two bets: Bet 1 doubled the money placed on it if a participant's score were higher than the other. Bet 2 doubled the money placed on it if a participant's score was lower than the other. In the event that their score was identical to that of the other, then the tie would be broken based on whose answer to the tiebreaker question was more accurate. The tiebreaker question was: "*How many minutes does it take light from the sun to reach the earth?*"[2] Participants were reminded that they did not have to gamble; they could be guaranteed to make $4 by placing $2 on each of the two bets, since one of them was guaranteed to win.[3]

After they had answered all these questions, participants were shown the correct answers to the quiz, shown the score of the other/typical person, given their monetary payoffs, thanked, and dismissed.

Results

Manipulation check. As expected, scores on the easy quiz were indeed higher (M = 8.53) than scores on the difficult quiz (M = 1.39), F (1, 176) = 1340, $p < .0001$, $\eta^2 = .88$.

Pre-test priors. Participants' estimates of the probabilities that they and others would receive each of the 11 possible scores are shown in Figure 2. As Figure 2 shows, participants reported holding similar priors for self and for others. Our theory has little useful to say about these priors and where they might come from. The real contribution of our differential information theory is a clearer understanding of how priors are updated in the presence of new information.

Post-test posteriors. One of the key predictions of our theory is that posterior beliefs about others will regress more toward the prior than will beliefs about self. This ought to show up as a significant difficulty x target interaction effect, because estimated performance will go up more for self than others on the easy quiz and it will go down more for self than for others on the difficult test. In order to test this prediction, we first computed post-test weighted estimated scores by multiplying each of the eleven scores by the probability assigned to it, and then summing these. We then submitted these estimates to a 2 (difficulty) x 2 (salience of other) x (2) (target: self vs. other) between-subjects mixed ANOVA, with repeated measures on target.

Of course, after having taken the test, people who took the easy quiz estimated that both they and others would obtain higher scores (M = 8.15) than did those who took the difficult quiz (M = 3.07), and this is reflected in a main between-subjects effect of quiz difficulty, F (1, 251) = 3280, $p < .001$, $\eta^2 = .74$.

The main effect of target is also significant, F (1, 251) = 15.83, $p < .001$, $\eta^2 = .06$. This effect reflects the fact that, on average, people estimated that they would score worse (M = 5.36) than would others (M = 5.73). While this effect is inconsistent with self-enhancement theories, it is readily explainable by differential regression. It is a direct result of the fact that people overestimated others more on the difficult quiz (*Mean overestimate* = 2.09) than they underestimated them on the easy quiz (*Mean underestimate* = 1.08), t (253) = 3.74, $p < .001$, η^2 =

.05. Since people's priors were fairly high to begin with, there was simply more room for regressive estimates of others to produce overestimates of performance on the difficult quiz.

As predicted, the two-way difficulty x target interaction emerged as significant, F (1, 251) = 39.26, $p < .001$, $\eta^2 = .14$. This interaction describes the fact that post-test estimates of performance are more extreme for self than for others: People estimated that they had done better (M = 8.26) than others (M = 8.05) on the easy quiz, t (123) = 1.90, $p = .06$, $\eta^2 = .03$, and worse (M = 2.62) than others (M = 3.54) on the difficult quiz, t (130) = -6.39, $p < .001$, $\eta^2 = .24$. This pattern is shown in Figure 3. This figure shows the distinctive pattern predicted by our theory: that people's beliefs about themselves are regressive, but that their beliefs about others are even more regressive. The differential weighting explanation cannot account for these results because it is silent regarding how people form individual assessments of self and other.

Furthermore, consistent with the idea that salience of the other influences beliefs, the three-way difficulty x target x salience interaction is significant, F (1, 251) = 3.93, $p = .049$, $\eta^2 = .02$. This interaction describes the fact that the difficulty x target interaction is stronger among those comparing themselves with the typical other, F (1, 120) = 30.62, $p < .001$, $\eta^2 = .20$, than it is among those comparing themselves with the person sitting next to them, F (1, 131) = 10.20, $p = .002$, $\eta^2 = .07$.

Where do evaluations about performance come from? Differential weighting theories are vague regarding exactly how these evaluations are made, but the claim that individual evaluations of self and other are similar suggests they arise through similar processes. By contrast, our theory makes a set of clear and specific predictions regarding the factors that influence beliefs about self and others: Belief should be based on both priors and actual performance. However, beliefs about others should be influenced far less heavily by actual performance than should beliefs about self. In order to test these predictions, we conducted two regressions using post-test weighted estimated scores for self and other as dependent variables, the results of which appear in Table 1.

Consistent with our theory, participants' own quiz performances exert a stronger influence on their post-test estimates of *self* (B = .72, SE = .06, t = 12.27, $p = 2.01 \times 10^{-27}$) than on post-test estimates of *others* (B = .36, SE = .06, t = 5.60, $p = 5.61 \times 10^{-8}$). When participants were estimating their own scores, they had useful information. They relied heavily on their own scores, but their pre-test priors were also a significant influence. When estimating the scores of others, the regression results suggest participants relied less heavily on their own experiences and instead tried to account for the ease or difficulty of the task—hence the significance of quiz difficulty.

The significant effects of pre-test priors for estimations of both self and others suggest, consistent with our theory, that people's priors affect their subsequent judgments. They updated from these priors using new information, and since the information they had (their own quiz performances) was more useful for estimating self than others, this information was weighted more heavily when estimating their own scores than when estimating others' scores.

This updating process is perfectly sensible, but it has interesting and non-intuitive consequences which are illustrated in Figure 4. Like the hypothetical data in Figure 1, Figure 4 shows that when the task was easier than expected, people underestimated their own scores but underestimated the scores of others even more so, leading them to believe that they were better than others. When the task was harder than expected, people overestimated their own scores but overestimated the scores of others even more so, leading them to believe that they were worse than others.

Comparative judgments. At this point, it should not be surprising that our direct measures of comparative judgments are quite consistent with the individual assessments. For instance, the correlation between the direct comparison and the computed difference between own and others' scores is .71. Furthermore, the various direct comparative judgments are roughly consistent with each other; participants also bet more on the easy quiz (M = \$2.63) than on the difficult quiz (M = \$1.89), t (253) = 4.44, $p < .001$, $\eta^2 = .07$. See Table 2.

Our theory holds that direct comparative judgments arise from the individual assessments. If this is so, then including the computed difference between self and other as a covariate in the omnibus ANOVA should decrease the effect of difficulty. Indeed, the inclusion of this measure decreases the effect size of the difficulty manipulation from $\eta^2 = .15$ ($p < .0001$) to $\eta^2 = .04$ ($p = .002$). Nevertheless, it is important to note that this computed difference is not redundant with the direct comparative judgment—the η^2 value indicates that the covariate accounts for only 44% of the variance in direct comparative judgments, and the effect of difficulty remains significant.

Reconciliation with prior results. If differential regressiveness in individual evaluations is so important, then why did Chambers and Windschitl (2004) conclude that “empirical findings do not suggest that [differential regression] plays a major role” in BTA and WTA effects (p. 828)? Chambers and Windschitl base this claim on findings such as those from Windschitl et al.’s (2003) fifth experiment, in which the effect of difficulty on direct comparisons was larger ($\eta^2 = .14$) than its effect on the computed difference between assessments of self and others ($\eta^2 = .01$). By contrast, in our experiment, the effect of difficulty on direct comparisons ($\eta^2 = .14$) was more similar to its effect on computed differences ($\eta^2 = .13$).

We suspect that the discrepancy between their result and ours is due to three important differences between our methods. First, our participants are evaluating something—prior trivia quiz performance—that is objective, specific, and clear. Burson and Klayman (2005) have shown that differential weighting plays a larger role when performance measures are vague or anticipatory than when they are clear and retrospective. Second, our participants' estimates of these specific real performances are made on objective unambiguous scales—that is, number correct out of ten. Subjective verbally-anchored scales (such when performance is rated on a scale that ranges from "*poor*" to "*excellent*") promote conflation between individual and relative evaluation (Biernat et al., 1997; Mussweiler & Strack, 2000). And this conflation can contribute to findings of differential weighting (Moore, in press). Third, we examined our participants' comparative and individual judgments on the same issue—test scores—and we find consistency between the two. Prior studies, such as Windschitl et al.'s fifth experiment, measure different issues. The individual assessments measured perceived knowledgability on trivia topics and the direct comparisons were the probability of winning a trivia contest; it ought not to be surprising that these are less consistent.

However, even in the present data, we *do* find evidence of some differential weighting, even on our direct measures of comparative judgment. When self and other are weighted optimally, the weighting on other is roughly 80% the size of the weighting on self (due to the greater variance in estimates of self). By contrast, actual weightings vary from 74% in the case of the direct comparison to 39% in the case of the verbally-labeled 1 to 7 rating scale. These results suggest that differential regressiveness accounts for the majority of BTA and WTA effects on unambiguous measures but less so on vague, subjective measures. On subjective measures, differential weighting explanations appear to hold more promise—but question vagueness appears to moderate this effect.

Discussion

The results of the first experiment are consistent with key predictions made by our differential information explanation. Participants' beliefs about the performance of others regressed toward pre-test priors more so than did their beliefs about their own performances. The resulting differential regressiveness accounts for the majority of the BTA and WTA effects we observe. If the results of Experiment 1 seem obvious and unsurprising, it is probably because our theory is so consistent with normative rationality. Yet we are the first to offer this simple and sensible explanation for BTA and WTA effects. Prior explanations have centered on egocentric overweighting of self-assessments, which our results suggest play a small role relative to differential regressiveness. This simple explanation has some counterintuitive implications, more of which we test in Experiments 2 and 3.

We should explicitly note a departure in our methodology from the paradigms traditionally used to study BTA effects and comparative optimism. It has been common for researchers to ask participants about tasks that occur outside the laboratory, such as their driving abilities or their friendliness toward others. While there are clear advantages in studying events and abilities that matter in everyday life, traditional approaches have two distinct shortcomings. First, because it is usually difficult if not impossible for researchers to obtain accurate objective measures of actual performance in these tasks, the traditional approach cannot tell us whether self-reports of performance by self and others are regressive with respect to actual performance. The three studies in this paper, on the other hand, all provide objective measures of actual performance, thereby allowing us to examine the relationship between realty and participants' estimates of self and others. This is obviously crucial for testing our theory. Second, because it is unclear what the correct standards for measuring performance on real-life tasks such as driving or friendliness, researchers are forced to use vague subjective measures of performance. As we have seen, these measures are noisier than objective measures and subject to conflation with related constructs.

EXPERIMENT 2: GUESSING WEIGHTS

Our theory posits a key role for the effect of information in comparative judgment. BTA and WTA effects should be magnified by high quality information about one's own performance, but they should be eliminated or reversed by high quality information about others' performance. To see how the effects could be reversed, imagine a situation in which neither of two students obtained any information about their professor's grade on her own paper, but each learned how the professor had graded the other one. In this situation, our theory would hold that the "self estimate" and "other estimate" lines in Figure 1 should switch places. In other words, if the professor was a lenient grader, each student would learn that her rival had gotten stunningly high marks, leading her to fear that she could not have measured up. And had the professor been a tough grader, each would learn that her rival had been graded badly, making her confident that she would do better. Because each had accurate information about the other, their estimates of others were less regressive than their estimates of self. As consequence, these students would believe that they were worse than others when the grading was lenient and better than others when the grading was tough.

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 employing a task in which participants do not know how well they have done after they have completed it. Experiment 2 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, or neither.

Our theory's predictions for this information manipulation distinguish it clearly from rival theories. Self-enhancement theories would predict that motives toward self-enhancement would exist in all conditions, but evidence suggests that self-enhancement exerts its strongest effect when people have clear information about others and their own performances are ambiguous (Klein, 2001). Differential weighting theories would predict that information about others would affect comparative judgments primarily by making others salient and increasing the weight put on estimates of their performance. They have nothing to say regarding information's effect on individual estimates of performance or the implicit comparative judgment they represent.

Method

Participants. Two hundred and fifty-one students at two large private U.S. universities participated in exchange for payment. All were given \$4 to bet on their own performances; results of these bets yielded an average payment of \$4.92 (SD = \$2.84). The experiment lasted approximately 15 minutes. Fifty-six percent of participants were female.

Design. The experiment had a 2 (easy vs. difficult) x 3 (feedback about self vs. feedback about other vs. no feedback) 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 web site that randomly assigned participants to conditions and led them through the experiment. 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 were shown a series of 10 photographs and had to guess how much each person in each picture weighed. Those in the easy 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 (for a similar manipulation of difficulty, see Burson et al., 2006). The 10 photographs appeared in a different random order for each participant.

After they estimated the 10 weights, roughly two-thirds of participants received feedback about performance. In all conditions, feedback was truthful. One third of participants were informed of their own scores (out of 10) on the test. One third of 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.[4] This set of previous test-takers scored an average of 8.76 using the easy criterion and 1.27 using the difficult criterion.[5] One third of participants received no feedback about performance.

Participants were then asked to make estimates of relative and absolute individual performance. In order to rule out idiosyncratic effects of question order, half the participants answered the comparative performance question first and half answered the individual performance questions first.

Comparative judgments. Participants were asked to directly 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*." This was the direct comparative judgment.

Estimates of individual performance. Participants were asked to make estimates of individual 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 questions appeared was counterbalanced.

Bets. Participants were then asked how they wanted to bet their $4. As in Experiment 1, they had two bets on which they had to wager all their money: "*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."

Results and Discussion

Omnibus test. We first conducted an omnibus 2 (difficulty) x 3 (feedback) ANOVA using participants' direct comparative judgments. Our key prediction, the moderating effect of feedback on BTA and WTA effects, would appear as a significant difficulty x feedback interaction effect, in which feedback would moderate the effect of difficulty. Consistent with this prediction, the difficulty x feedback interaction is significant, F (2, 244) = 11.27, $p < .001$, $\eta^2 = .09$. The two main effects, both qualified by this interaction, also emerge as significant. The main effect of difficulty is significant, F (1, 244) = 13.85, $p < .001$, $\eta^2 = .05$. And the main effect of feedback is significant, F (1, 244) = 4.91, $p = .008$, $\eta^2 = .04$.

Below we seek to clarify the exact form of the key interaction, and further explore the results relevant to testing our theory.

BTA and WTA effects in direct comparisons. Because they obtained no useful feedback about their own performances, we will use the reports of the participants in the no-feedback as a measure of baseline beliefs, given the experience of taking the test but without the benefit of feedback about performance. Among these participants, there was no significant difference in direct comparative judgments between the easy (M = 1.00) and difficult (M = .31) conditions, t (76) = 1.24, p = .22, $\eta^2 = .02$. They did report believing, on average, that they had scored .68 points above average, and this is significantly different from zero by one-sample t-test, t (77) = 2.43, p = .018, $\eta^2 = .07$, but this BTA belief was not moderated by test difficulty. The lack of an effect of test difficulty makes sense given that the experience of this test, unlike the test used in Experiment 1, provides little to no useful information for judging one's performance.

Relative to this baseline, our theory predicts stronger BTA and WTA effects among those who only got feedback about their own performances, and weaker BTA and WTA effects among those who only got feedback about others. In order to simplify the analysis of these dual effects we used participants' direct comparative judgments and took the negative of participants' judgments for those participants in the difficult condition. Larger positive values of this BTA/WTA index measure would indicate stronger BTA effects in the easy condition and WTA effects in the difficult condition, whereas negative numbers would indicate the reverse.

The three feedback conditions differ significantly from each other on this BTA/WTA index, F (2, 247) = 11.76, $p < .001$, $\eta^2 = .09$. In the no-feedback baseline condition, the BTA/WTA index was .40. As our theory would predict, the self-feedback condition showed stronger BTA and WTA effects ($M = 1.27$), and a planned contrast reveals this difference to be significant, t (247) = 2.61, $p = .010$. Furthermore, the other-feedback condition showed a significant reversal of BTA and WTA effects relative to the baseline ($M = -.29$), t (247) = -2.11, $p = .036$. See Figure 5.

BTA and WTA effects in individual assessments. Table 3 presents the means for individual assessments in the three experimental conditions. Consistent with our theory and with the results discussed above, when these individual assessments are subject to a 2 (difficulty) x 3 (feedback) x (2) (self vs. other) mixed ANOVA, the three-way interaction emerges as significant, F (2, 245) = 5.54, $p = .004$, $\eta^2 = .04$. This three-way interaction results from the fact that the standard effect (BTA beliefs on easy tasks and WTA beliefs on difficult tasks) occurred most strongly among participants who got feedback about their own scores. This standard effect manifests itself in a two-way difficulty x target interaction, which is highly significant among those who got self-feedback, F (1, 79) = 17.59, $p < .001$, $\eta^2 = .18$. This interaction effect is more modest among those who got no feedback, F (1, 76) = 8.63, $p = .004$, $\eta^2 = .10$. And it is eradicated among those who were given other-feedback, F (1, 90) = .002, $p = .97$.

Our theory would predict a reversal of the standard BTA and WTA effects when people get better information about others than themselves, assuming that no additional effects were operating. While comparative judgments in this condition do represent a significant reversal from the baseline in the no-feedback condition, the fact remains that, as Table 3 shows, those who get other-feedback in the easy condition do not report believing that they are worse than others. This fact is probably due to the additional effect, demonstrated by Klein (2001; Klein, Monin, Steers-Wentzell, & Buckingham, 2006), for self-enhancement motivations to exert the strongest effect on self-estimates when people have clear information about others' performances but lack information about their own performances. Nevertheless, the differential weighting explanation cannot account for the patter of BTA and WTA effects we observe in individual assessments, nor can it account for the close parallel between direct comparisons and individual assessments.

EXPERMENT 3: OTHERS' TRIVIA

Experiment 2 helps to disentangle egocentric effects of evaluating the self from the effects of having better information about the self. However, while self-enhancement motives cannot account for the WTA effects observed in Experiment 2, they do appear to have affected the results in at least one cell of our design. We could more clearly observe the effect of information if we minimized the role of self-enhancement motives. Because our theory applies just the same to the evaluation of any two individuals (or groups, for that matter) about whose performances the evaluator has differential information, Experiment 3 seeks to replicate the results of Experiment 2 in the evaluation of two randomly selected individuals, thereby minimizing the role of motivational effects on judgment. Again, our theory predicts that the target about whom individuals have the most information will be the one whom they will predict to perform better on easy tasks and worse on difficult tasks. In contrast, the differential weighting theory would predict that the known target will be weighted more heavily, but the theory is silent on the question of how people make individual assessments and cannot account for BTA and WTA effects in individual assessments.

Method

Participants. The 113 volunteers were primarily undergraduate students from a large private U.S. university. Females constituted 56% of participants.

Procedure. Instructions began: "*In February of 2003, twenty-four students (both undergraduates and graduate students) signed up to participate in a trivia contest. Today you will be asked to consider two of those people, Person K and Person J. These are two actual people who competed against each other in the trivia contest. We will ask you to predict who won, and bet on your prediction. This research is studying people's ability to make accurate inferences about others based on only selective information*." Indeed, the stimuli were taken from another study in which participants took either an easy or a difficult ten-item trivia quiz and also wrote an autobiographical paragraph about themselves in answer to the question, "What makes me unique?" (Moore & Kim, Experiment 4).

Participants in the present study were then told that they had $4 that they would have to bet on whether the focal target's score had exceeded that of the other person. We will refer the non-focal person as the referent. We counterbalanced whether we referred to the focal target as

Person K or Person J. For each participant, a different Person J and Person K were selected randomly from the set of 24 previous participants, with the constraint that both target and referent had taken the same quiz.

Participants were always provided with the autobiographical paragraph written by one of the two people. The provision of such paragraphs has been shown to increase the salience of and focusing on that individual (Moore & Kim, 2003), but the paragraphs are devoid of information useful to estimating the person's trivia quiz performance. The individual whose paragraph a participant received thereby became the focal target. Half of participants also received a copy of the target's completed quiz. The other half of participants saw a copy of the referent's completed quiz. The order in which participants encountered either the paragraph or the completed quiz was counterbalanced. Note that the completed quiz was not graded, so participants did not have perfect information about the score of the person whose test they saw. They were not provided with the correct answers and so could not be sure if the answers they saw were correct.

Design. The design, then, was a 2 (difficulty) x 2 (information: participant sees completed quiz of target vs. referent) between-subjects design.

Dependent measures. Participants first estimated the number of questions the target had answered correctly and made a direct comparative judgment that asked them to estimate how many more questions the target had answered correctly than had the referent. Participants were then invited to bet on whether the target would beat the referent. Again, participants had to allocate all of their $4 between Bet 1 (which paid off if the target beat the referent) and Bet 2 (which paid of if the referent beat the target). Then participants estimated the number of questions that the referent had answered correctly.

Finally, we included questions designed to serve as manipulation checks of information and focusing. For the target and the referent, participants were asked, "How much useful information did you have for estimating trivia quiz performance?" They were given a 7-point scale on which to respond, with endpoints labeled "*no useful information*" and "*a great deal of useful information*." Also, for the target and referent participants were asked, "*To what extent did you put yourself in their perspective?*" They were given a 7-point scale on which to respond, with endpoints labeled, "*not at all*" and "*very much so*."

Results

Manipulation checks. The manipulation check confirms that participants thought they had more useful information about the person whose test they saw than about the other person. We tested this using a 2 (see test of target vs. referent) x (2) (target vs. referent) mixed ANOVA. This ANOVA produced the expected 2-way interaction, F (1, 111) = 114.24, $p < .001$, $\eta^2 = .51$. When participants saw the target's test, they reported that they had better information about the target (M = 4.72) than the referent (M = 1.46), t (56) = 14.02, $p < .001$. And when participants saw the referent's test, they reported that they had worse information about the target (M = 3.13) than the referent (M = 4.48), t (55) = -3.71, $p < .001$. Note that this interaction effect qualifies two main effects: Participants reported that they had more useful information about the target (M = 3.93) than the referent (M = 2.96), F (1, 111) = 19.44, $\eta^2 = .15$. This suggests that they thought that seeing the target's autobiographical paragraph was somehow useful in estimating their quiz performance. And they reported that they had more useful information in total when they saw the referent's test (M = 3.80) than when they saw the target's test (M = 3.09), F (1, 111) = 9.89, p = .002, $\eta^2 = .08$.

As for the test of our manipulation to get participants to focus on the target, the manipulation check suggests that it was effective. We submitted participants' answers to the perspective-taking questions to a 2 (see test of target vs. referent) x (2) (target vs. referent) mixed ANOVA. As expected, participants reported that they took the perspective of the target (M = 4.61) more than that of the referent (M = 3.07), F (1, 111) = 44.96, $p < .001$, $\eta^2 = .29$. Surprisingly, the interaction effect also emerges as significant, F (1, 111) = 15.14, $p < .001$, η^2 = .12. This interaction effect reflects the fact that participants took the perspective of the target more when they saw the target's test (M = 4.82) than when they did not (M = 4.39). And they reported taking the referent's perspective more when they had the referent's test (M = 3.75) than when they did not (M = 2.40). Because they tended to report taking both the target's and the referent's perspective more when they saw the referent's test (M = 4.07) than when they saw the target's test (M = 3.08), the main effect whose test they saw also emerges as marginally significant, F (1, 111) = 3.77, p = .055, η^2 = .03.

The results on the manipulation checks suggest that the manipulations were not perfectly clean. People reported believing that the target's autobiographical paragraph was actually useful for estimating the person's trivia quiz sore. And people reported that seeing the referent's test led them to focus on that person and take their perspective. Nevertheless, the manipulations' effects were largely as intended. And to the degree that they were not, it would only have made it harder for us to find our effects.

BTA and WTA effects in individual assessments. Among those who saw only the target's test, we expected to replicate the standard BTA and WTA effects. On the other hand, among those who saw only the referent's test, we expected to find a reversal of the standard effect. Indeed, this reversal shows up as a significant difficulty x information interaction in the omnibus

2 (difficulty: difficult vs. easy) x 2 (information: participant saw completed quiz of target vs. referent) ANOVA performed on the computed difference between participants' estimates of the target's and the referent's scores, F (1, 109) = 13.24, $p < .001$, $\eta^2 = .11$.[6] See Figure 6 and Table 4. Among those who saw only the target's test, in the easy quiz condition they predicted that the target would score better than the referent on the easy quiz (M = .66) but worse than the referent on the difficult quiz (M = -.25), and a planned contrast reveals this difference to be significant, t (109) = 2.05, p = .043. Those who saw only the referent's test predicted that the target would score worse than the referent on the easy (M = -1.21) but better on the difficult quiz (M = .16), t (109) = -3.09, p = .003.

One striking feature of Figure 6 is how much smaller the effects are in the difficult than in the easy condition. This is because of an interesting unanticipated difference we created when we gave participants others' ungraded quizzes. In the easy condition, participants had a good sense of which items were correct and which were incorrect, because most of them knew the answers themselves. But in the difficult condition, because most participants did not know the correct answers to the questions, such as, "What is the largest moon of Saturn?" participants had more trouble determining others' scores. Indeed, in response to the manipulation check question, "How much useful information did you have for estimating trivia quiz performance?" those in the difficult condition rated their information as significantly less useful (M = 3.16) than did those in the easy condition (M = 3.73), p = .013. As a result, participants' estimates of scores were more regressive in the difficult than in the easy condition. On average, those in the difficult condition overestimated scores by 2.57, whereas those in the easy condition underestimated scores by .72.

BTA and WTA effects in direct comparative judgments. Again, direct comparative judgments are consistent with individual assessments. The 2 (difficulty) x 2 (information) ANOVA reveals a significant interaction effect for both direct comparative judgments, F (109) = 10.40, $p = .002$, $\eta^2 = .09$, and for bets, F (109) = 11.42, $p = .006$, $\eta^2 = .07$. The pattern of these interactions matches that for individual assessments, and neither of the main effects is significant for either direct comparisons or for bets.

Causes of differential regressiveness. To what degree does the differential regressiveness that produces BTA and WTA effects result from differences in information versus focusing on the salient target? We constructed an index of regressiveness by subtracting participants' estimates of the target's score from the target's actual score in the easy condition and the reverse in the difficult condition. A 2 (information: saw test of target vs. referent) x (2) (target vs. referent) mixed ANOVA reveals that our manipulation of focusing had no significant effect on the regressiveness of estimates, F (1, 111) = .02, $p = .88$, $\eta^2 < .01$. Estimates of the target were no less regressive ($M = 1.65$) than were estimates of the referent ($M = 1.62$). However, as our differential information theory would predict, the target x information interaction is significant, F (1, 111) = 14.26, $p < .001$, $\eta^2 = .11$. This effect results from the fact that participants' estimates of both the target and the referent became less regressive when they saw their tests ($M = 1.28$) than when they didn't ($M = 1.99$).

Discussion

The results of Experiment 3 replicate those of Experiment 2 without the complicating feature of self-enhancement motives. Our theory does not distinguish between self-assessments and assessments of others. This is not to say that we believe that self-assessments are identical to assessments of others or that we think our theory explains everything. Voluminous evidence makes it clear that there are important differences between evaluations of self and others. It goes without saying that our theory does not explain all features of comparative judgment—there is no theory that can. Rather, our theory and experiments isolate one distinguishing feature of self-judgments, superior information, and shows that this feature can account for BTA and WTA effects more parsimoniously than can any of its rival theories.

GENERAL DISCUSSION

The results of the three experiments we present are consistent with the predictions of our theory. The most basic and fundamental prediction we make is that information about someone's performance makes estimates of that performance less regressive. This prediction is supported by results from all three experiments. Because of this, an individual who is known to have performed poorly is judged to be worse than others, whereas an individual who is known to have performed well is judged better than others. Again, this prediction is supported by all three experiments. In particular, Experiment 1 tracked the effect of information on prior beliefs and showed that information about performance led people to update their prior beliefs so as to produce BTA and WTA effects. The effect of information is so strong that it can produce a reversal of the standard BTA and WTA effects when the person making the comparative judgment has more information about the referent than the target, as the results of Experiments 2 and 3 show.

There are no other existing theories that can account for the results we present. Theories of self-enhancement or self-serving bias cannot explain the WTA effects evident in each of our three experiments. Differential weighting explanations cannot account for the substantial BTA and WTA effects we observe in individual assessments, because differential weighting explanations are silent regarding the origins of individual assessments and they do not predict

BTA or WTA effects in individual assessments. Our results support differential regressiveness over differential weighting because we obtain BTA and WTA effects when we compare individual assessments in addition to when we examine comparative judgments.

Our differential information explanation suggests that BTA and WTA effects result from sensible rules of inference. The theory presented in this paper may be so consistent with normative principles that it seems unsurprising. If it does, then we have achieved our aim: to show that there is a simpler explanation for BTA and WTA effects than those that have been offered in previously published work. We have also used this normative explanation to posit more intriguing implications such as: (1) the opposite effects of difficulty on overestimation of one's individual score and overestimation of one's performance relative to others; and (2) the reversal of the usual BTA and WTA effects we find in Experiments 2 and 3.

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. Indeed, it would be difficult to argue that rational people should not exhibit these effects. Even so, it is clear that BTA and WTA effects are biases, with potentially profound effects on behavior. They could lead to excessive entrepreneurial entry in industries that are perceived as easy, such as restaurants, bars, and clothing retail (Moore & Cain, in press; Moore, Oesch, & Zietsma, in press). They could lead to excessive confidence in contests where all competitors are strong, such as in the final tournament of a sports league (Radzevick & Moore, 2006). And they could lead to excessively high costs of conflict (such as strikes, court battles, or wars) when each side is strong.

Regression Effects

The more difficult the task, the more people tend to overestimate their individual 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 performances will underestimate excellent performances and overestimate terrible performances (Burson et al., 2006; Erev et al., 1994; Krueger & Mueller, 2002). Here, we demonstrate that such regression effects can also account for BTA and WTA effects in comparative judgment. Our results indicate that BTA and WTA effects are attributable in large part to greater regressiveness in estimates of others than of the self. As a result, people estimate the behavior of others as less extreme than their own, resulting in BTA effects on easy tasks and WTA effects on difficult tasks. In our three experiments, better information about the target than the referent led to differential regression.

Our results also found evidence of differential weighting. However, its causal role in BTA and WTA effects appears modest when compared to the effect of differential regression. The effect of differential weighting appeared strongest with subjective verbally-anchored rating scales. Clearer, less ambiguous direct comparative measures, on the other hand, appeared to be less prone to bias and produced smaller differences between direct comparisons and individual assessments. And more importantly, our differential information explanation provides something that the differential weighting explanation does not: an explanation for how people make individual estimates of self and other that can explain BTA and WTA effects implicit in individual assessments.

The differential information explanation offers a parsimonious theory of the underlying psychological processes that produce both BTA and WTA biases in comparative judgment. This explanation is not inconsistent with focalism accounts for BTA and WTA effects, especially the version of the focalism account that highlights the informational privilege often held by focal individuals. Focusing can also influence the degree to which people project self-assessments on to others, as the results of Experiment 1 show. It is worth noting, however, that focusing manipulations can be strong enough to induce people to apply what information they have about performance selectively to the target of their focus. For instance, in Moore and Kim's (2003) fourth experiment, differential regression accounted for 66% of BTA and WTA effects on bets by those focusing on themselves but only 16% of the BTA and WTA effects among those focusing on the opponent, about whom they had little information. Furthermore, as our results make clear, differential information is not the only cause of differential regression.

Overconfidence

A great deal of research has examined the phenomenon of overconfidence. Many researchers refer to overestimates of one's own individual performance as overconfidence (Erev et al., 1994; Griffin, Dunning, & Ross, 1990; Polivy & Herman, 2002; Soll & Klayman, 2004). Others refer to overestimates of one's performance relative to others as overconfidence (Bazerman & Neale, 1982; Camerer & Lovallo, 1999; Cooper et al., 1988; Neale & Bazerman, 1985). The results presented here highlight the importance of distinguishing between overestimation of individual performance and overestimates of comparative performance (see also Larrick, Burson, & Soll, in press). 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; informational influences on absolute and relative confidence judgments can run in opposite directions. Differential regression can reconcile the apparent inconsistency between evidence on the hard/easy effect and evidence of BTA and WTA effects.

Conclusion

In this paper, we have focused on the value of the differential information explanation to help account for BTA and WTA effects such as: people believing that they are better than others on easy tasks, that they are more likely than others to engage in common behaviors, that they are more likely than others to display common traits, that they are more likely than others to experience common events, and that they are less likely than others to experience rare events. But the fact that people usually have better information about themselves and their ingroups than about other people and other groups has far broader explanatory power, as Fiedler (1996) has argued. Differential information leads people to see the self as more complex than others (Sande et al., 1988) and ingroups as more complex and heterogeneous than outgroups (Jones, Wood, & Quattrone, 1981). More familiarity with ingroups can produce illusory correlations in which common behaviors are seen as more typical of the ingroup than the outgroup and rare behaviors are seen as more typical of the outgroup (Hamilton & Gifford, 1976; Sanbonmatsu, Sherman, & Hamilton, 1987). Greater information about focal hypotheses than rival hypotheses produces prior beliefs that can lead evidence against the focal hypotheses to be evaluated more critically than information favoring it, leading to confirmatory biases in hypothesis testing (Lord, Ross, & Lepper, 1979). Differential information has profound affects on human judgment that go beyond the effects on which we have focused in this paper.

The theory we present highlights the importance of understanding how people use information in the assessment of self and others. Psychologists have been critical of theories, such as economic theories, that make unrealistic assumptions about people's abilities and their knowledge. But we must be careful not to make similar errors ourselves, by assuming either that people have perfect information about themselves or about others. Clearly, they do not. We can understand human judgments better when we understand their origins. That means understanding the assimilation and processing of information on the way to making a judgment. The evidence in this paper suggests that simple and sensible processes in information acquisition and processing can produce surprisingly complicated and counterintuitive results.

References

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

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.

Babcock, L., & Olson, C. (1992). The causes of impasses in labor disputes. *Industrial Relations, 31*, 348-360.

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., & Kobrynowicz, D. (1997). Simultaneous assimilation and contrast effects in judgments of self and others. *Journal of Personality and Social Psychology, 73*(2), 254-269.

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

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

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

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

Chambers, J. R., & Suls, J. (in press). The role of egocentrism and focalism in the emotion intensity bias. *Journal of Experimental Social Psychology*.

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.

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior, 6*, 151-155.

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

Dawes, R. M. (1990). The potential nonfalsity of the false consensus effect. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn.* (pp. 179-199). Chicago, IL, USA: University of Chicago Press.

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

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and business. *Psychological Science in the Public Interest, 5*(3), 69-106.

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.

Epley, N., & Dunning, D. (2006). The mixed blessings of self-knowledge in behavioral prediction: Enhanced discrimination but exacerbated bias. *Personality and Social Psychology Bulletin, 32*(5), 641-655.

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

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

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

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

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.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory Correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*, 392-407.

Healy, P. J., & Moore, D. A. (2006). *Bayesian overconfidence: Theory and experiments*. Unpublished manuscript.

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

Jones, E. E., Wood, G. C., & Quattrone, G. A. (1981). Perceived variability of personal characteristics in ingroups and outgroups: The role of knowledge and evaluation. *Personality and Social Psychology Bulletin, 7*, 523-528.

Karniol, R. (2003). Egocentrism versus protocentrism: The status of self in social prediction. *Psychological Review, 110*(3), 564-580.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*(1), 32-37.

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.

Klein, W. M. P., Monin, M. M., Steers-Wentzell, K. L., & Buckingham, J. T. (2006). Effects of Standards on Self-Enhancing Interpretations of Ambiguous Social Comparison Information. *Basic and Applied Social Psychology, 28*(1), 65-79.

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

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

Krueger, J. I., & Clement, R. W. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology, 67*(4), 596-610.

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

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*(2), 221-232.

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

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

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

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

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

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization. *Journal of Personality and Social Psychology, 37*(11), 2098-2109.

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.

Malmendier, U., & Tate, G. (2004). *Who makes acquisitions? CEO overconfidence and the market's reaction*. Unpublished manuscript, Berkeley, CA.

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

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.

Miller, D. T., & McFarland, C. (1987). Pluralistic ignorance: When similarity is interpreted as dissimilarity. *Journal of Personality and Social Psychology, 53*(2), 298-305.

Moore, D. A. (2006). *When good = better than average*. Unpublished manuscript, Pittsburgh.

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

Moore, D. A., & Cain, D. M. (in press). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior & Human Decision Processes*.

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

Moore, D. A., & Klein, W. M. P. (2006). *Objective standards matter too much: The use and abuse of absolute and comparative performance feedback in absolute and comparative judgments and decisions*. Unpublished manuscript.

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

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

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110*(3), 472-489.

Mussweiler, T., & Strack, F. (2000). The "relative self": Informational and judgmental consequences of comparative self-evaluation. *Journal of Personality and Social Psychology, 79*(1), 23-38.

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.

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

Polivy, J., & Herman, C. P. (2002). If at first you don't succeed: False homes of self-change. *American Psychologist, 57*(9), 677-689.

Radzevick, J. R., & Moore, D. A. (2006). *For the love of the game? Betting, prediction, and myopic bias in athletic competition*. Unpublished manuscript, Pittsburgh.

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.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology, 37*, 322-336.

Sanbonmatsu, D. M., Shavitt, S., & Gibson, B. D. (1994). Salience, set size, and illusory correlation: Making moderate assumptions about extreme targets. *Journal of Personality and Social Psychology, 66*(6), 1020-1033.

Sanbonmatsu, D. M., Shavitt, S., Sherman, S. J., & Roskos-Ewoldsen, D. R. (1987). Illusory correlation in the perception of performance by self or a salient other. *Journal of Experimental Social Psychology, 23*(6), 518-543.

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

Sande, G. N., Goethals, G. R., & Radloff, C. E. (1988). Perceiving one's own traits and others': The multifaceted self. *Journal of Personality and Social Psychology, 54*(1), 13-20.

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

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

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193-210.

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.

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.

Footnotes

[1] Although our theory is consistent with Bayesian rationality (for a formal mathematical development, see Healy & Moore, 2006) we do not assume that people update their beliefs in perfect Baysian fashion, only that they observe the crudest standard for coherence: that their beliefs tend to lie between their prior expectations and actual outcomes.

[2] Correct answer: 8.3 minutes.

[3] We employed this complicated bet structure for two reasons. First, it allowed participants to bet on being either worse than or better than others. Second, it reduced the confound between risk preferences and beliefs about comparative performance, since participants had to bet all their money. Had we just asked them how much they wanted to bet on being better than the other, they could have chosen not to bet either because (1) they thought they were worse or (2) because they were risk averse.

[4] These 330 other participants were part of another study, the results of which are reported elsewhere (Moore & Klein, 2006).

[5] Because comparison with the group average with skewed distributions will lead the majority of people to be above average on the simple task (with the negatively skewed distribution) and below average on the difficult task (with the positively skewed distribution), we ought to expect to find BTA and WTA effects. However, our interest in the present study is the effect of feedback manipulations on the size of these BTA and WTA effects. This interest is not compromised by the presence of baseline BTA and WTA effects.

[6] This same 2-way interaction shows up just the same if, instead of using the computed target-referent difference as the dependent measure, we use direct comparative judgments, F (1, 109) =

10.40, $p = .002$, $\eta^2 = .09$, or bets (on the target beating the referent), $F(1, 109) = 7.97$, $p = .006$, $\eta^2 = .07$. It also appears as a significant 3-way interaction if, instead of using the computed target-referent difference, we include both as a repeated measure in a 2 x 2 x (2) mixed ANOVA, $F(1, 109) = 13.25$, $p < .001$, $\eta^2 = .11$.

Appendix

Trivia questions used in Experiment 1.

Easy	Difficult
1. Berlin is the capital of what country?	1. What is the capital of Azerbaijan?
2. How many hours are in one day?	2. How many nanoseconds are there in a second?
3. The United States shares the longest unguarded border in the world with its neighbor to the north. What is the name of that country?	3. Before becoming a Canadian province, Saskatchewan was part of what other entity?
4. How many meters are there in a kilometer?	4. How many feet are there in a mile?
5. What is the name of the prophet of the Islamic faith, born in the city of Mecca in the year 571?	5. What is the most popular first name in the world?
6. Which US state, in which the bumper sticker "Don't mess with ____" is popular, is known as the Lone Star state?	6. Which US state instituted the nation's first mandatory seat-belt law in 1984?
7. The Golden Gate Bridge is located in which North American city?	7. Which North American city has the following subway stops: Kendall Square, Central Square, and Porter Square?
8. In what European country is Dutch spoken?	8. How many countries were members of the European Union as of June 2003?
9. The Roman god of war gives his name to the "Red Planet," the fourth planet from the Sun in our solar system. What is his name?	9. Who was the Greek God of War?
10. Baghdad is the capital of what country?	10. Who ruled Iraq before Saddam Hussein?
Tiebreaker question:	How long does it take light from the sun to reach the surface of the earth?

Answers—Easy: (1) Germany (2) 24 (3) Canada (4) 1000 (5) Mohammed (6) Texas (7) San Francisco (8) Holland/the Netherlands (9) Mars (10) Iraq. Difficult: (1) Baku (2) 1 billion or 10^9 (3) The Northwest Territories (4) 5280 (5) Mohammed (6) New York (7) Cambridge, MA (8) 15 (9) Ares (10) Ahmad Hassan al-Bakr. Tiebreaker: 8.3 minutes.

Table 1

Regressions predicting post-test score estimates for self and other, Experiment 1. (Standard errors in parentheses.)

Model 1 predicting post-test beliefs about own performance		Model 2 predicting post-test beliefs about other's performance	
Independent variable	Unstandardized *B* coefficient	Independent variable	Unstandardized *B* coefficient
Pre-test estimated score for self	.22*** (.05)	Pre-test estimated score for other	.28*** (.06)
Own actual score	.72*** (.06)	Own actual score	.36*** (.07)
Difficult quiz dummy	-.63 (.44)	Difficult quiz dummy	-2.02*** (.49)
R^2	.84***	R^2	.72***

*** $p < .001$

Table 2

Six different measures of comparative judgment compared, Experiment 1. The second column shows the effect size attributable to the difficulty manipulation. The third and fourth columns show the results of path analyses predicting the comparative judgment on that row using participants self-reported post-test estimates of performance by self and other.

	Simple vs. Difficult	Regression results		Correlations with	
Self-reported Comparative Judgment	Effect size (η^2)	β (Self)	β (Other)	Actual Percentile	Actual Score
Estimated difference between self and others (direct comparison)	.15***	1.37***	-1.01***	.31***	.46***
Bet	.07***	.98***	-.64***	.38***	.38***
Estimated win likelihood	.24***	1.13***	-.65***	.30***	.53***
Estimated percentile rank	.24***	1.01***	-.42***	.41***	.59***
Relative rating (1-7 scale)	.33***	1.09***	-.42***	.30***	.67***

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

Table 3

Table showing estimated individual scores for self and others, Experiment 2. Standard deviations appear in parentheses.

	Difficult		Easy	
	Self	Others	Self	Others
No feedback	5.28 (2.19)	5.44 (1.59)	7.33 (2.07)	6.31 (1.80)
Self Feedback	2.94 (2.03)	3.66 (1.3)	8.00 (2.37)	6.85 (1.86)
Other feedback	3.31 (1.70)	2.85 (1.73)	8.38 (1.53)	7.90 (1.01)

Table 4

Table showing estimated individual scores for target and referent. Standard deviations appear in parentheses, Experiment 3.

	Difficult		Easy	
	Target	Referent	Target	Referent
See target's test	3.61 (1.66)	3.86 (1.63)	9.48 (0.74)	8.83 (1.1)
See referent's test	4.68 (1.76)	4.52 (1.67)	8.32 (2.11)	9.54 (0.79)

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

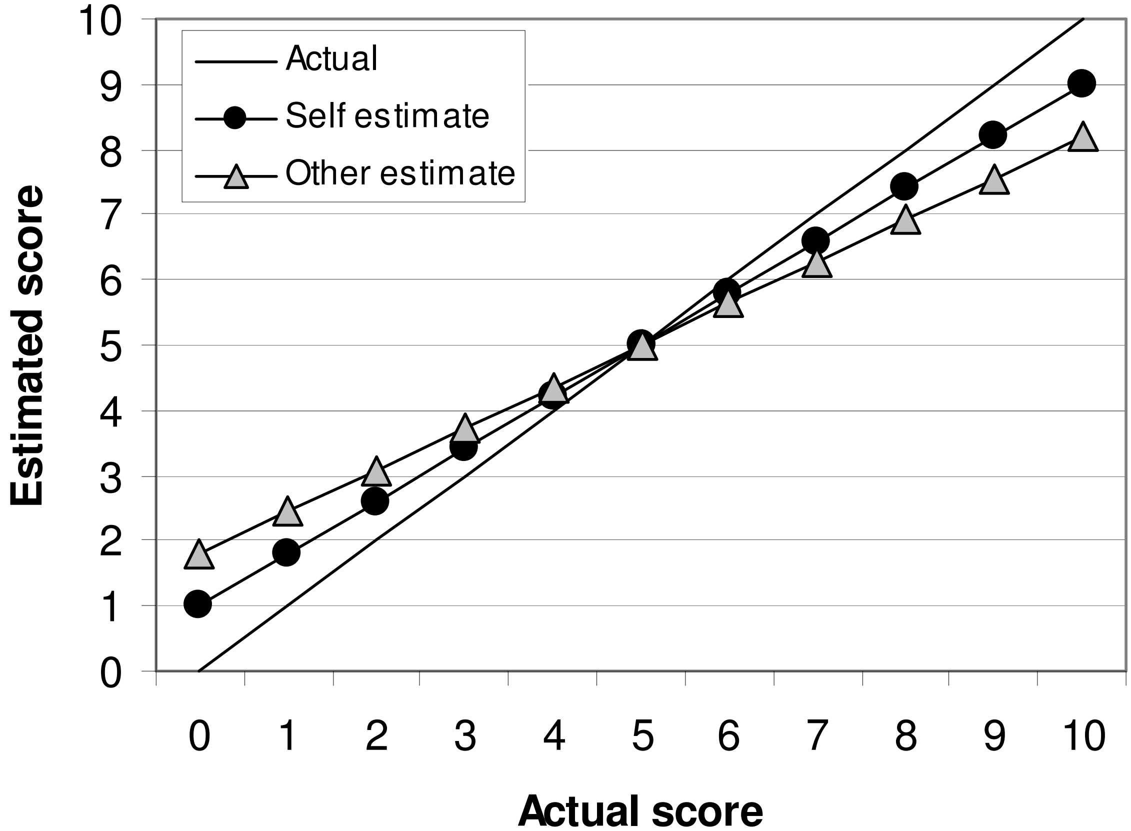


Figure 2. Pre-test estimated probabilities of obtaining each possible score, for self and for others, Experiment 1.

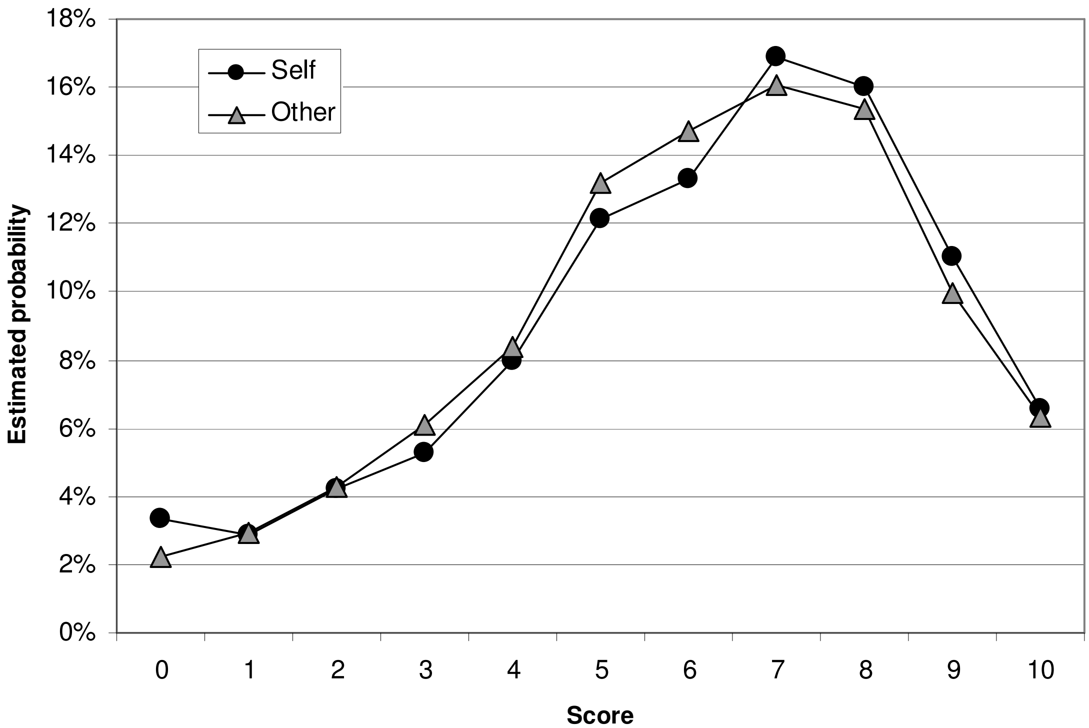
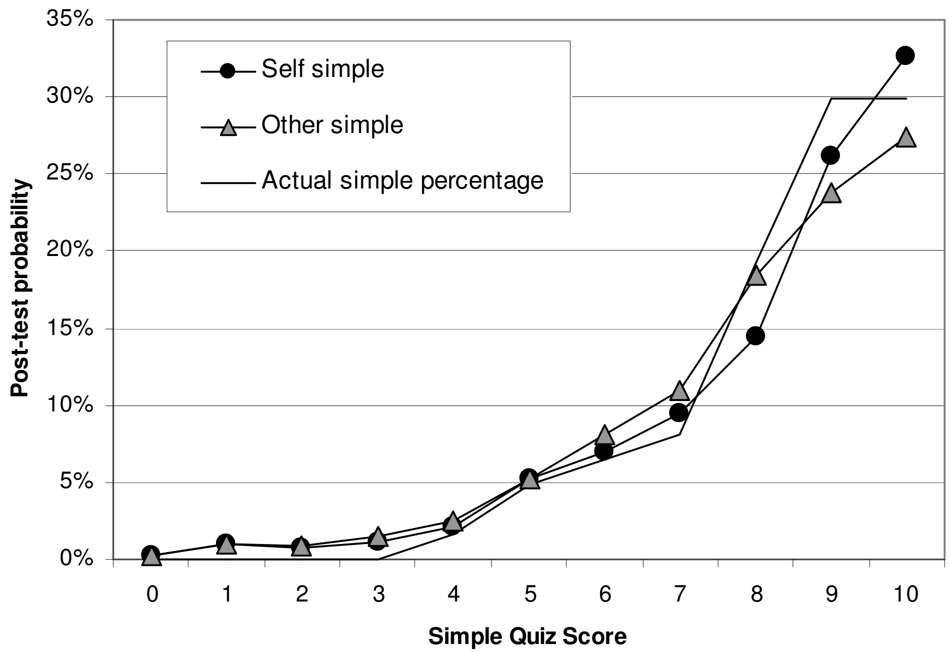


Figure 3. Post-test estimated probabilities of having obtained each possible score, for self and for others (Experiment 1). The top panel shows only those participants who took the simple quiz. The bottom panel shows those who took the difficult quiz.



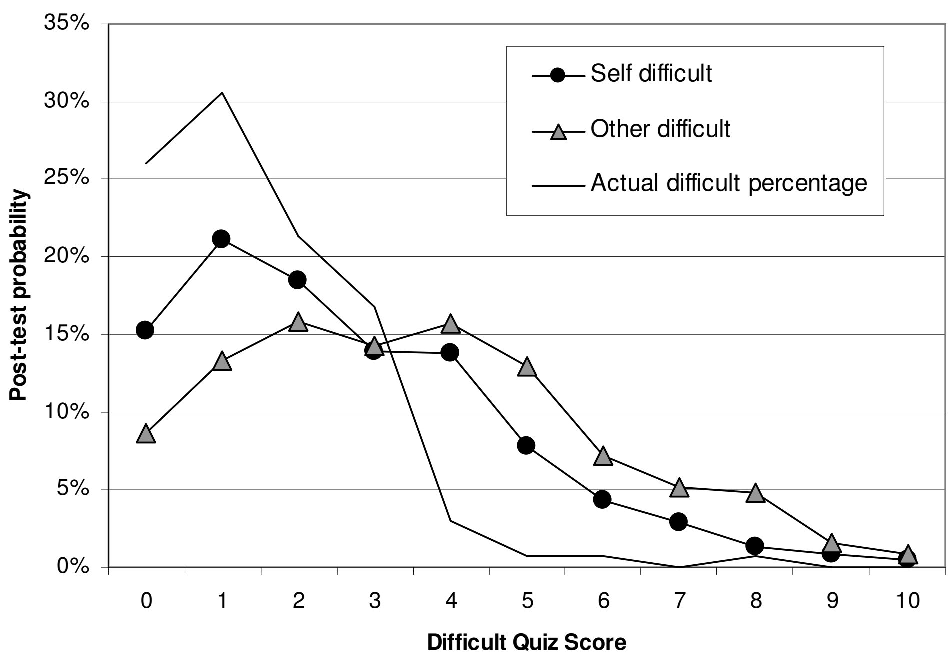


Figure 4. Participants' self-reported beliefs about own and others' performance, as a function of their own actual scores, Experiment 1. (Error bars show standard errors.)

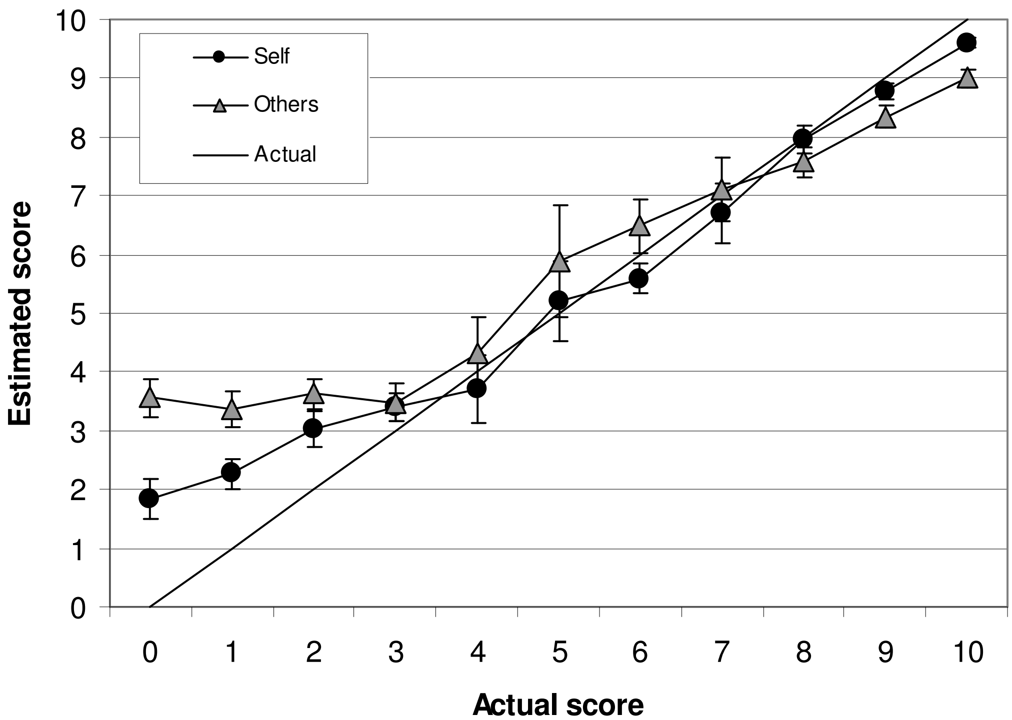


Figure 5. Participants' estimates of their performances relative to the average, as a function of task difficulty and feedback about their own performances, Experiment 2. A rating of 1 indicates that participants in that condition believed they scored 1 point above average. Negative scores indicate that people thought they were below average. Error bars show standard errors.

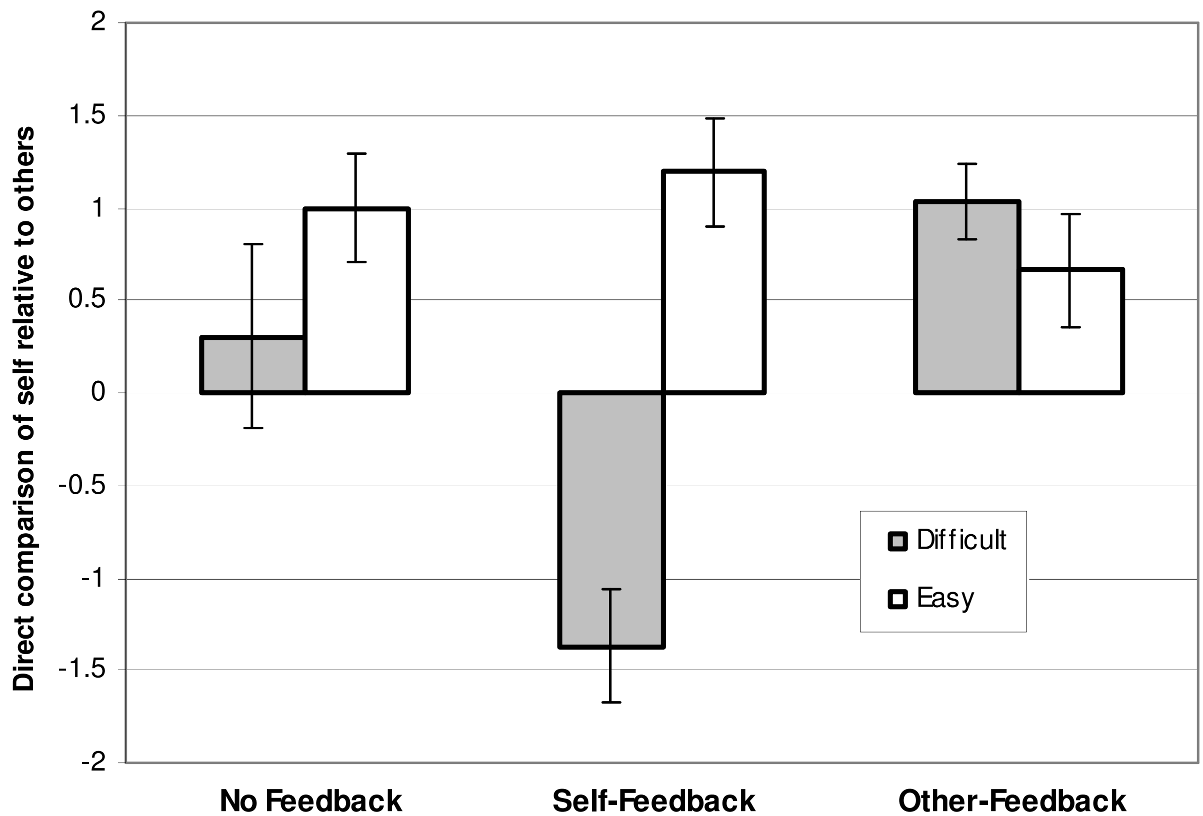


Figure 6. Participants' estimates of the performance of the target relative to that of the referent, Experiment 3. Error bars show standard errors.