Journal of Personality and Social Psychology
2002, Vol. 82, No. 2, 180–188

0022-3514/02/$5.00 DOI: 10.1037//0022-3514.82.2.180

Unskilled, Unaware, or Both? The Better-Than-Average Heuristic and Statistical Regression Predict Errors in Estimates of Own Performance

Joachim Krueger and Ross A. Mueller
Brown University

People who score low on a performance test overestimate their own performance relative to others, whereas high scorers slightly underestimate their own performance. J. Kruger and D. Dunning (1999) attributed these asymmetric errors to differences in metacognitive skill. A replication study showed no evidence for mediation effects for any of several candidate variables. Asymmetric errors were expected because of statistical regression and the general better-than-average (BTA) heuristic. Consistent with this parsimonious model, errors were no longer asymmetric when either regression or the BTA effect was statistically removed. In fact, high rather than low performers were more error prone in that they were more likely to neglect their own estimates of the performance of others when predicting how they themselves performed relative to the group.

Demonstrations of cognitive–perceptual biases have been central to social–psychological research since the breakdown of normative attribution theories in the 1970s. Ordinary social perceivers have been shown to reason egocentrically and to be insensitive to the rules of scientific inference. At the same time, they are said to be overconfident in the accuracy of their own judgments (Gilovich, Griffin, & Kahneman, 2002; Nisbett & Ross, 1980). Chief among the social–perceptual biases is the "better-than-average" (BTA) effect. Most people believe that they are better and that they do better than the average person (Alicke, 1985; Brown, 1986; Krueger, 1998b). The BTA effect emerges in a variety of judgment domains, such as personality descriptions, risk perceptions, and, with the exception of very difficult tasks, expectations of performance (Kruger, 1999). Although researchers debate its adaptive value (e.g., Asendorpf & Ostendorf, 1998), most agree that the BTA effect reflects irrational thinking because "it is logically impossible for most people to be better than the average person" (Taylor & Brown, 1988, p. 195). When the BTA effect is found as a group phenomenon, it is tempting to conclude that it characterizes people in general. But such a conclusion would be rash. Of the many people who believe themselves to be better than average, many actually are (Krueger, 1998a). The question then becomes: Who is biased and why?

Kruger and Dunning (1999) recently showed that poor performers greatly overestimate their own performance, whereas high performers slightly underestimate theirs. According to Kruger and Dunning, high performers possess metacognitive skills that enable them to understand their own abilities. Poor performers, in contrast, not only "reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the ability to realize it" (p. 1121). In four studies, participants completed a test (of grammar, logic, or humor appreciation), and they provided percentile estimates for their own performance relative to the performance of others. Participants were then grouped in quartiles according to their own actual percentiles, and for the bottom and the top quartile participants, estimated percentiles were compared with their corresponding actual percentiles by *t* tests. Although Kruger and Dunning recognized that regression effects virtually guarantee asymmetric estimation errors, they believed "that the overestimation we observed was more psychological than artifactual. For one, if regression alone were to blame for our results, then the magnitude of miscalibration among the bottom quartile would be comparable with that of the top quartile" (p. 1124).

We suggest that in conjunction with an overall BTA effect, statistical regression can account for asymmetric estimation errors. As long as the correlation between the predictor variable (actual percentiles) and the criterion variable (estimated percentiles) is imperfect, "the variance of our predictions should never be larger than that of the criterion we seek to predict (Never, not just hardly ever)" (Goldberg, 1991, p. 181). With the slope of the regression line being smaller than 1, we expected poor performers to overestimate their own percentiles and high performers to underestimate theirs. When there is an overall BTA effect, poor performers make larger errors than high performers. Any increase in the BTA effect raises the regression line and thereby sharpens the asymmetry in the errors.

The relationship between the height of sons and the height of their fathers is a classic illustration of regression (Galton, 1886). When successive generations become taller—as has been the case in the western world during the twentieth century—the shortest fathers have sons who are much taller than they themselves are, whereas the tallest fathers have sons who are a little shorter than they themselves are. Mediator variables are not necessary to ex-

Joachim Krueger and Ross A. Mueller, Department of Psychology, Brown University.

Ross Mueller is now at the Fuller Theological Seminary, Pasadena, California.

This research was conducted as part of Ross Mueller's honors thesis, which was awarded the Harold Schlosberg Memorial Premium in Psychology. We are indebted to Melissa Acevedo, Jill Krueger, and Judith Schrier for their thoughtful comments on a draft of this article.

Correspondence may be addressed to Joachim Krueger, Department of Psychology, Brown University, Box 1853, Providence, Rhode Island 02912. E-mail: Joachim_Krueger@Brown.edu

plain this asymmetry. Indeed, it is hard to imagine what such mediators might be (Krueger, 2000b).

In the domain of human performance, the role of mediators is more plausible. Kruger and Dunning (1999) defined metacognitive skill as "the ability to know how well one is performing, when one is likely to be accurate in judgment, and when one is likely to be in error" (p. 1121). To measure this skill, they asked participants to predict which of their responses were correct and which were incorrect. Then, they computed the sum of "the number of questions each participant accurately identified as correct or incorrect" (p. 1128). When this mediator is controlled, the correlation between the predictor (i.e., performance) and the criterion (i.e., estimation errors) should be reduced.[1] We expected this result because even if participants have no insight into their successes and failures across test items, high performers make more correct predictions than poor performers. The reason is the BTA heuristic, which itself does not indicate discriminatory metacognitive skill, but only a general optimistic bias. Consider a high performer who solves 80% of the test items. As an optimist, this person may expect to be correct 60% of the time. Lacking metacognitive skill, however, this person does not know which responses are correct. Nevertheless, most predictions of success and failure are correct by chance. Successes are identified with a probability of .6, and failures are identified with a probability of .4, so that the total percentage correct is 56% (i.e., $[(.8 \times .6) + (.2 \times .4)] \times 100$). A poor performer, who is equally optimistic and who also lacks metacognitive skill, is correct 44% of the time.

Kruger and Dunning (1999) also considered the role of social projection. Rather than treating projection as a metacognitive skill, they viewed it as an impediment faced only by the top performers, who underestimate their performance because they "fall prey to the false consensus effect" (p. 1126). Expecting others to do as well as they themselves do, they fail to realize how much better they did than others. From the perspective of regression, the underestimation is expected of the top performers because values on the predictor variable (actual percentiles) rise faster than do values on the predicted variable (estimated percentiles). Because the beneficial effects of projection on accuracy are well documented (Dawes, 1989; Krueger, 1998c), we considered projection a potential metacognitive skill and examined it as such.

We tested the mediational model implied by the metacognitive hypothesis using several unbiased indexes of metacognitive skill as well as the confounded measure of percent correct. We expected that only the latter would yield significant mediation effects. Next, we derived and tested competing predictions. According to the regression–BTA hypothesis, asymmetric estimation errors require both regression and the BTA effect. When either one of these group effects is controlled, the asymmetry should disappear. Moreover, only the regression–BTA hypothesis predicted symmetric errors when performance measures were corrected for unreliability. Finally, only the metacognitive hypothesis predicted that poor performers would be most likely to neglect the performance of others when estimating how well they did relative to the group.

Method

Following Kruger and Dunning (1999, Study 3), we selected test items from the National Teacher Examination preparation guide (Bobrow et al., 1989). Sixty-two volunteers completed 50 test items. Twenty items were selected to construct a difficult test (M correct = 29%, SD = 14.84), and 20 items were selected to construct an easy test (M correct = 70%, SD = 16.08; see Appendix for sample items). Thus, the easy test was similar to the test used by Kruger and Dunning (M = 67.5%, p. 1125).

Twenty-five male and 35 female undergraduate students (average age = 19.63 years) participated individually as part of a research requirement. A program written in Superlab (Haxby, Parasuraman, Lalonde, & Abboud, 1993) presented the test items, the follow-up questions, and the rating scales on a computer. The program also stored all responses. Participants were randomly assigned to the difficult and the easy test condition. After responding to a test item, they rated their degree of confidence in the accuracy of their response on a scale ranging from 1 (*not confident*) to 8 (*highly confident*). Using the same scale, they rated how confident they were that a majority of students at their university could answer the item correctly. Finally, they estimated the number of questions that they had answered correctly, and they predicted their own percentile rank of the test.

Results

Manipulation Checks

The estimated and actual numbers of correct responses were submitted to an analysis of variance (ANOVA), in which the difficulty of the test varied between participants. The effects of type of measure, $F(1, 58) = 74.54$, $p < .001$, and of test difficulty were statistically significant, $F(1, 58) = 101.29$, $p < .001$, as was the interaction between the two variables, $F(1, 58) = 54.94$, $p < .001$. Actual scores were lower on the difficult test ($M = 5.53$, $SD = 2.18$) than on the easy test ($M = 13.90$, $SD = 2.34$), $F(1, 58) = 209.17$, $p < .001$, and estimates were also lower for the difficult test ($M = 12.10$, $SD = 3.26$) than for the easy test ($M = 14.40$, $SD = 2.46$), $F(1, 58) = 15.81$, $p < .001$. Participants overestimated their test scores only when the test was difficult, $F(1, 29) = 127.01$, $p < .001$, but not when it was easy ($F < 1$).

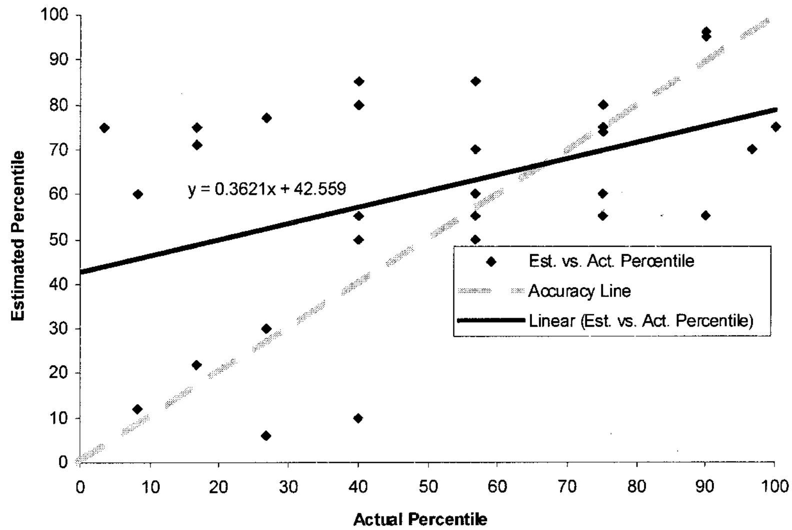
Replication

Mean estimated percentiles exceeded the 50% mark for both the difficult test ($M = 61.27$, $SD = 24.08$), $t(29) = 2.56$, $p < .05$, and for the easy test ($M = 68.77$, $SD = 19.41$), $t(29) = 3.18$, $p < .01$. The BTA effect was somewhat larger for the easy test, although this difference was not statistically significant, $t(58) = 1.33$, $p < .10$. Across participants, actual and estimated percentiles were moderately correlated (difficult: $r = .44$; easy: $r = .14$). Thus, the sufficient sources of the error asymmetry were present: The BTA effect occurred on the group level, and actual and estimated performance percentiles were positively, yet imperfectly, correlated. In Figure 1 (top panel: difficult test; bottom panel: easy test), estimated percentiles are plotted against actual percentiles. The regression lines show that poor performance involved large overestimation errors, whereas high performance involved small underestimation errors.

Regression and Mediation

Actual percentiles were negatively correlated with estimation errors (i.e., estimated − actual percentiles) for both the difficult

[1] In Kruger and Dunning's (1999) main analysis, the percent correct measure predicted estimation errors when performance was controlled. The partialed variable was the predictor instead of the mediator. The mediator variable was controlled only when poor performers' estimation errors were predicted from an experimental manipulation of training (vs. no training).



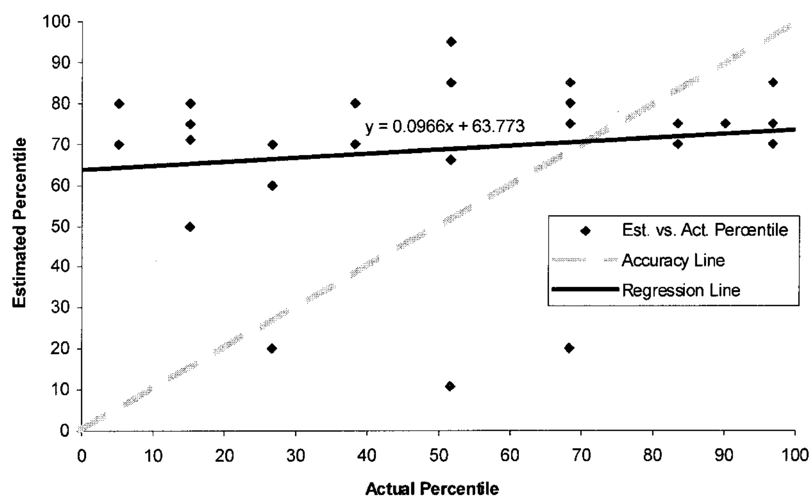


Figure 1. Estimated versus actual percentiles on a difficult (top) and on an easy (bottom) test. Est. = estimated; Act. = actual.

($r = -.65$) and the easy test ($r = -.80$). These correlations were fully determined by the correlations between the predictor (actual percentiles) and the criterion variables (estimated percentiles) and their variances.[2] As McNemar (1969) noted, "no [difference] scores ever need to be calculated" (p. 177). Against this background of statistical dependency, the metacognitive hypothesis postulated a significant role for mediator variables. For mediation to occur, a mediator variable must to be correlated with the predictor. If, as expected, this correlation is positive, a negative correlation between the mediator and the criterion is likely because the criterion variable is a difference score involving the predictor itself (i.e., estimation errors = estimated − actual percentiles). Nevertheless, we can ask whether control of the mediator variable reduces the negative correlation between predictor and criterion.

To test the mediational hypothesis comprehensively, we considered five different measures. The first measure was percent correct. We dichotomized confidence ratings, assuming that rat-

[2] With x representing actual percentiles and y representing estimated percentiles, the correlation between x and y − x is

$$r_{x,y-x} = \frac{r_{xy}s_y - s_x}{\sqrt{s_x^2 + s_y^2 - 2r_{xy}s_xs_y}}.$$

Any decrease in the variance of y increases the regression effect. Because they are forced to range from 0% to 100%, actual percentiles are more variable (s_x = 29% for each test) than estimated percentiles (s_y = 24% for difficult and 19% for easy).

ings from 1 to 4 indicated expectations of failure and that ratings from 5 to 8 indicated expectations of success. The percent correct measure was correlated with actual percentiles ($r = .58$ for difficult and .85 for easy), and was thus negatively correlated with estimation errors ($r = -.60$ for difficult and $-.65$ for easy).

The second measure was an adjusted percent correct score, which we computed by subtracting the correct responses that would occur by chance from the total percent correct score. This measure was not correlated with actual percentiles ($r = .27$ for difficult and $-.02$ for easy), and it was negatively correlated with estimation errors ($r = -.08$ for difficult and $-.01$ for easy).

The third measure indexed the ability to discriminate between one's own successes and failures across test items. This skill was expressed by the correlation between self-related confidence ratings and actual outcomes (correct vs. incorrect). On average (after *r-Z-r* transformation), discriminative skill was low for the difficult test ($M = .12$) and medium for the easy test ($M = .40$). Skill was positively related to actual percentiles ($r = .48$ and .18) and negatively related to errors ($r = -.31$ and $-.04$).

The fourth measure indexed the ability to discriminate between the successes and failures of other test takers. Other-related confidence was correlated with the actual success rate of others (M for difficult = .14, M for easy = .28). This measure was also positively related to actual percentiles ($r = .19$ and .30, for difficult and easy, respectively) and negatively related to errors ($r = -.20$ and $-.15$, for difficult and easy, respectively).

The fifth measure indexed social projection as the correlation between confidence ratings in the quality of one's own performance and confidence regarding the performance of others. Projection scores (M for difficult = .74, M for easy = .76) were positively related to actual percentiles ($r = .18$ and .20, for difficult and easy, respectively) and negatively related to estimation errors ($r = -.32$ and $-.11$, for difficult and easy, respectively).

To test the mediational hypothesis, we computed the correlations between actual percentiles and estimation errors while controlling for each mediator variable, one at a time. The degree to which the partial correlations were smaller than their corresponding zero-order correlations was tested for significance with modified Sobel tests (Kenny, Kashy, & Bolger, 1998). As can be seen in Table 1, the percent correct measure partially mediated the correlation between actual percentiles and estimation errors ($z = 2.69$, $p < .01$, and $z = 3.93$, $p < .0001$, for the difficult and the easy test, respectively). This effect was expected because this measure confounded metacognition with actual performance. More importantly, seven of the eight tests involving unbiased mediators changed the correlation by .01 or less. The one test that involved a change of .05 (for own discrimination on the difficult test) was not significant ($z = 1.40$, $p = .16$).[3]

Comparisons between the two test conditions further supported the regression–BTA hypothesis. According to the metacognitive hypothesis, larger average values on the mediator variables should have been associated with lower correlations between estimated and actual percentiles. However, for the unbiased mediator variables, the opposite was the case.

Table 1
Mediated and Unmediated Correlations Between Actual Percentiles and Estimation Errors

	Test	
Correlation	Difficult	Easy
Zero-order correlation	.65	−.80
Mediator controlled		
Percent correct	−.48	−.63
Adjusted percent correct	−.65	−.81
Own discrimination	−.60	−.81
Other discrimination	−.64	−.81
Projection	−.64	−.80

Competing Predictions

According to the regression–BTA hypothesis, estimation errors should no longer be asymmetric when either regression or the BTA effect is removed. According to the metacognitive hypothesis, however, poor performers might continue to show disproportionately large overestimation errors. We controlled the regression effect for the 8 poorest performers by estimating the regression equation using only the data of the remaining participants. We then predicted the estimated percentiles for the poorest performers under the assumption of linearity. For the difficult test, the average of the residual errors (i.e., estimated percentiles − predicted percentiles) was in the direction predicted by the metacognitive hypothesis, but small in size ($M = 7.79\%$, $d = .25$). For the easy test, there was no discernible residual error ($M = 1.04\%$, $d = .05$; both ts < 1).

We controlled the overall BTA effect (i.e., the average percentile estimate – 50%) by subtracting it from the estimation errors. To realize the computation of the corrected values, consider the difficult test. Bottom quartile participants overestimated their performance by 35%, whereas top quartile participants underestimated their performance by 10%. The sum of the two errors was 25% (i.e., 35% + [−10%]). When the BTA effect of 11% was subtracted from each error, the corrected error in the bottom quartile was 24%, whereas it was –21% in the top quartile. The sum of the corrected errors was 3%, which supported the regression–BTA hypothesis. Table 2 displays the data from both test conditions along with the data from Kruger and Dunning's (1999) studies. When averaged, the corrected error asymmetry was near zero ($M = 2\%$).

Analyses across studies further supported the regression–BTA hypothesis. As expected, the size of the error asymmetry covaried perfectly with the size of the regression effect (i.e., with $1 - r$). When the BTA effect was removed, this relationship disappeared.

Statistical Power

Most of the predictions generated by the metacognitive hypothesis call for the rejection of null hypotheses, suggesting that the

[3] Several additional indexes also all failed to yield mediation effects: (a) *Pr*, which is the difference between the hit rate (predicted successes divided by all successes) and the false-positive rate (incorrectly predicted success divided by all failures; Snodgrass & Corwin, 1988); (b) indexes involving absolute, as opposed to signed, estimation errors; and (c) projection as measured by the signed or unsigned differences between self-related and other-related confidence ratings.

Table 2
Data for Bottom and Top Quartile Performers From Six Studies

			Bottom quartile			Top quartile			Error asymmetry	
Study	*r*	BTA	Est%	Act%	Error	Est%	Act%	Error	Raw	−BTA
Present article										
1	.44	11	50	15	35	76	86	−10	25	3
2	.14	19	65	15	50	68	87	−19	31	−7
Kruger and Dunning (1999)										
1	.39	16	58	12	46	75	88[a]	−13	33	1
2	.09[a]	11	62	12	50	68	86	−18	32	10
3	.19	18	61	10	51	70	89	−19	32	−4
4	.40	11	53	13	40	79	90	−11	29	7
Correlation with *r* across studies		−.38	−.89	.26	−.86	.94	.27	.95	−.91	.19

Note. *r* = the correlation between estimated (Est%) and actual (Act%) percentiles across all participants; Better-than-average effect (BTA) = overall mean estimated percentile − 50%; error = Est% − Act%; raw error asymmetry (Raw) = error in bottom quartile + error in top quartile; corrected asymmetry is the sum of errors with BTA subtracted in each quartile.
[a] Values estimated from Kruger and Dunning's (1999) figures.

more parsimonious regression–BTA model is not true. The regression–BTA hypothesis cannot be confirmed with the traditional practices of significance testing; it can only be retained so long as no alternative model is backed by significant evidence (Krueger, 2001; Nickerson, 2000).

In the present research, only generous increases in statistical power would offer hope for some of the relevant comparisons to reach significance. The mediational analyses would require a sample size of about 100 to attain power levels of .3 to .6 depending on the reliability of the mediator and assuming that the path from the predictor variable to the mediator variable is at least .4 (Hoyle & Kenny, 1999). For tests of competing predictions, a small to medium effect size ($d = .25$) can be detected with a power of .5 with 88 (one-tailed) or 140 (two-tailed) participants (Cohen, 1988). However, finding significance in isolated tests would not relieve the burden of explaining why these tests and not any of the many others were significant. For these reasons, and because of its parsimony and its ability to explain most of the systematic variance in the estimation errors, we retain the regression–BTA hypothesis.

Unreliable Performance Measures

Thus far, we have assumed that actual percentiles are perfectly reliable measures of ability. As in any psychometric test, however, the present test scores involved both true variance and error variance (Feldt & Brennan, 1989). With repeated testing, high and low test scores regress toward the group average, and the magnitude of these regression effects is proportional to the size of the error variance and the extremity of the initial score (Campbell & Kenny, 1999). In the Kruger and Dunning (1999) paradigm, unreliable actual percentiles mean that the poorest performers are not as deficient as they seem and that the highest performers are not as able as they seem. When a test lacks reliability, repeated testing places different people in the lowest and in the highest performance quartiles. Judging from our own data, the chances that a particular participant would be found in the same extreme quartile again were slim ($p = .32$ for difficult and .49 for easy). Estimation errors derived from a single test are therefore inflated for members of the extreme groups. When test scores are corrected for unreliability, estimation errors become less variable, and their correlation with actual percentiles becomes attenuated.

An analysis of test reliability is necessary to separate systematic estimation errors from random errors (Klayman, Soll, González-Vallejo, & Barlas, 1999). To perform such an analysis, we split both the difficult and the easy tests into subtests by separating the odd and even numbered items. We then regressed estimation errors on actual percentiles in two different ways. In the same-test method, actual percentile scores from the same subtest were used to place participants on the *x*-axis reflecting their performance and to predict their estimation errors as displayed in the *y*-axis (see the thick regression lines in Figure 2). In the different-test method, actual percentile scores on one subtest were used as the predictor, whereas scores on the other subtest were used to compute the criterion (the thin lines). The reliability of the difficult test was so modest (Spearman-Brown $r = .17$) that the error asymmetry was reversed (Figure 2, top). The easy test, which was more reliable (Spearman-Brown $r = .56$), still showed a substantial decrease in the asymmetry (Figure 2, bottom).

Self-Focus

A person's performance relative to the group rises with increases in that person's own performance and with decreases in the performance of others. The calculation of actual percentiles guarantees the effects of both variables, but there is no such guarantee for estimated percentiles. Indeed, when estimating how well they themselves did relative to others, people rely mainly on their absolute judgments of themselves. Klar and Giladi (1999) proposed a self-focus hypothesis according to which people transform their absolute sense of success or failure into judgments of how they did relative to others. In doing so, they fail to adjust suffi-

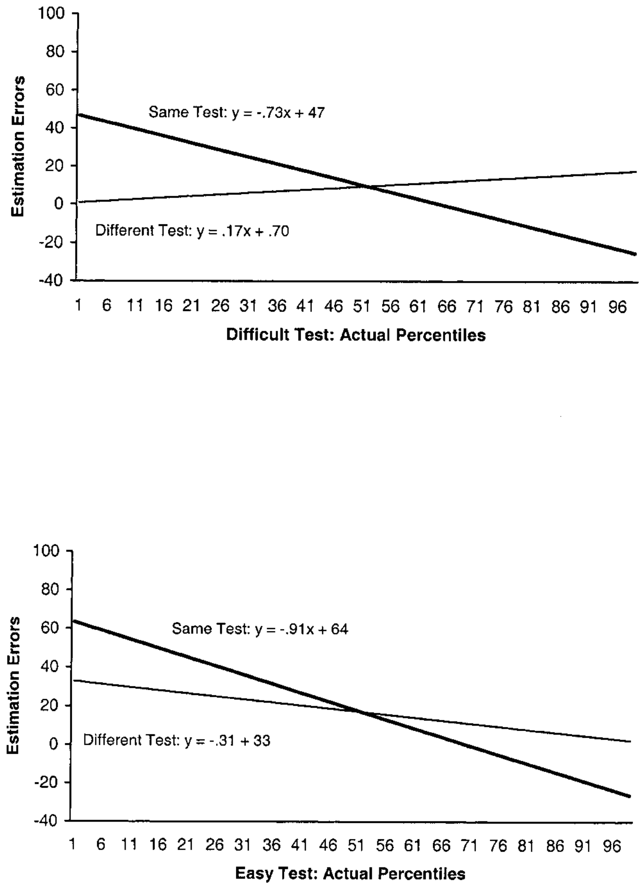


Figure 2. Regression of estimated errors on actual performance before (thick line) and after (thin line) correction for unreliability.

ciently for the effect of test difficulty on others (Kruger, 1999). Psychologically, self-focus appears to be grounded in the high accessibility and affective importance of self-related information (Clement & Krueger, 2000; Dunning & Hayes, 1997; Krueger & Stanke, 2001). Normatively, however, a selective focus on the self is problematic because it can lead to incoherent judgments. For example, self-focused participants who express low confidence in their own success might predict low performance percentiles even when they are less confident in the success of others.

Across participants, one would expect a positive correlation between estimated percentiles and self-related confidence and a negative correlation between estimated percentiles and other-related confidence. The self-focus hypothesis was supported in that the absolute magnitude of the former was greater than the magnitude of the latter for both the difficult test ($r = .51$ vs. $-.32$, $t = 1.39$, $p < .1$) and the easy test ($r = .40$ vs. $-.14$, $t = 2.62$, $p < .01$). Analysis of individual differences in self-focus presented a final opportunity to discriminate between the metacognitive hypothesis and the regression–BTA hypothesis. Only the former implied that poor performers would be less sensitive to social-comparison information than high performers would be (Kruger & Dunning, 1999, p. 1131). Thus, poor performers should be more erroneously self-focused. To test this hypothesis, we separated poor from high performers by median split and repeated the above analysis. Contrary to the metacognitive hypothesis, the correlations in Table 3 show that only the high performers neglected their own perceptions of how others were doing.

Discussion

Our theoretical analysis suggested that errors in the predictions of one's own performance can be explained by the regression of these predictions to an overall inflated mean. This interpretation is parsimonious; it does not require mediation by third variables, such as metacognitive insights into one's own problem-solving abilities. Our empirical analyses supported this view. None of the unbiased measures of metacognition mediated the relationship between actual percentiles and prediction errors. Tests that discriminated between the metacognitive hypothesis and the regression–BTA hypothesis favored the latter. The regression–BTA hypothesis also accounted for two findings considered anomalous under the metacognitive hypothesis. The first finding was that the top performers underestimated their percentiles; the second finding was that "although bottom-quartile participants accounted for the bulk of the above-average effects . . . there was also a slight tendency for the other quartiles to overestimate themselves . . . —a fact our metacognitive analysis cannot explain" (Kruger & Dunning, 1999, p. 1132). With regression to the mean and the overall BTA effect, one need not expect anything else.

Given the state of the correlational evidence, experimental work would be most informative if it manipulated metacognitive skill without altering competence. Then, one could ask whether changes in metacognition affect performance estimates. Kruger and Dunning (1999) conducted two experiments to "rule out the regression effect alternative" (p. 1128), but did not manipulate metacognitive skill directly. In Study 3 of their research, participants evaluated a set of completed tests before estimating their own performance again. Most participants increased their percentile estimates. The increase was significant only for the high performers, but it was not significantly larger than the increase among the poor performers. Thus, it remains unclear whether high and low performers reasoned differently. In Study 4, the authors manipulated only competence (i.e., the predictor variable). We agree with Kruger and Dunning that it is paradoxical to suggest "that the way to make incompetent individuals realize their own incompetence is to make them competent" (p. 1128). When improvements in competence are certain to beget improvements in metacognitive skills, the mediational role of those skills has little meaning. Some mediator

Table 3
Correlations Between Average Self-Related and Other-Related Confidence Ratings With Estimated Performance Percentile

	Test					
	Difficult			Easy		
Actual performance	Self		Others	Self		Others
Poor	.46		−.60	.39		−.26
High	.48	>	.08	.43	>	−.08

Note. The > sign indicates a statistically significant difference at $p < .05$ between the absolute size of the correlations.

variables can be manipulated directly. Shepperd (1993) found that poor performers, more than high performers, reported inflated Scholastic Assessment Test scores. When accuracy was rewarded, however, these distortions disappeared almost entirely. Contrary to the metacognitive hypothesis, poor performers were aware of their lack of ability and were motivated to disguise it.

Still, the metacognitive hypothesis is intuitively appealing. People are ready to assume that good traits as well as bad traits go together (Schneider, 1973). It is compelling to think that some people are smart, have accurate self-perceptions, and have metaintelligence too. This Platonic vision of coherent qualities focuses on distinctions among people, not on distinctions among their properties. In our view, however, human properties are diverse and poorly correlated with one another. Although perceptual, social, and academic skills form bundles (Stanovich & West, 1998), their qualitative and quantitative differences are too great to permit the extraction of a Spearman *g* factor of goodness (Krueger, 2000a). A final look at our data reveals this complexity. The correlations in Table 4 show that three different measures of self-enhancement cohered only in the difficult-test condition (top part of the table), and that three measures of metacognitive skill were empirically distinct regardless of test difficulty. If "the skills that engender competence in a particular domain are often the very same skills necessary to evaluate competence in that domain" (Kruger & Dunning, 1999, p. 1121), these variables should have been homogeneous within clusters and negatively correlated across clusters. If this were so, the question of statistical mediation would not even arise.

The metacognitive hypothesis assumes that the BTA effect is a mark of irrational thinking. Kruger and Dunning (1999) noted that "the tendency of the average person to believe he or she is above average defies the logic of descriptive statistics" (p. 1122). They also suggested that the people who show smaller BTA effects (i.e., the poor performers) reason more poorly than do the people who show larger BTA effects (i.e., the high performers). The inconsistency between these two positions may be resolved if we abandon the view that the psychological processes underlying the BTA heuristic must be defensive or distorted. Our concluding proposition is that the BTA effect can arise from rational reasoning under uncertainty.

Given the inevitability of imperfect estimation, the view that people, on average, ought to believe that they perform like the average person implies that only the average person is accurate. When the BTA heuristic comes into play, the intersection of the regression line with the accuracy line is displaced upward so that again only a few individuals are accurate. But if only few estimates are accurate regardless of the size of the BTA effect, self-enhancement by itself cannot constitute irrationality. Instead, the rationality of performance estimates can be assessed by asking how people make these estimates and how they evaluate the possible consequences.

Suppose a person chooses between predicting to be better than average and predicting to be worse than average. Either prediction can be true or false, creating four distinct events (see Figure 3). The hedonic value of each event depends on the valence of the outcome (success vs. failure) and on the valence of the expectation (optimism vs. pessimism; Shepperd, Ouellette, & Fernandez, 1996; Wedell & Parducci, 2000). A hit (positive self-verification) is pleasant because it is the conjunction of two positives. A false positive is less pleasant because disappointment follows optimism. In a miss, relief follows pessimism. A correct rejection (negative self-verification) is the conjunction of two negatives, although the resulting distress may be offset by the feeling of being right (Swann, 1984). To explain the BTA effect, it is only necessary to

Table 4
Correlations Among Measures of Bias and Measures of Metacognitive Skill

Measure	Self-enhancement 1	Self-enhancement 2	Self-enhancement 3	Metacognition 4	Metacognition 5	Metacognition 6
Difficult test						
Self-enhancement						
1. Estimated − actual percentile	—					
2. Estimated − actual score	.67**	—				
3. Confidence: self − other	.44*	.64**	—			
Metacognition						
4. Projection	−.32	−.11	.19	—		
5. Discrimination (self)	−.31	−.13	−.04	−.12	—	
6. Discrimination (other)	−.20	.06	−.06	.16	.17	—
Easy test						
Self-enhancement						
1. Estimated − actual percentile	—					
2. Estimated − actual score	.65**	—				
3. Confidence: self − other	.10	.18	—			
Metacognition						
4. Projection	−.11	−.20	.00	—		
5. Discrimination (self)	−.04	−.06	.13	.13	—	
6. Discrimination (other)	−.15	−.09	.13	.41*	.24	—

Note. $df = 28$.
* $p < .05$. ** $p < .001$.

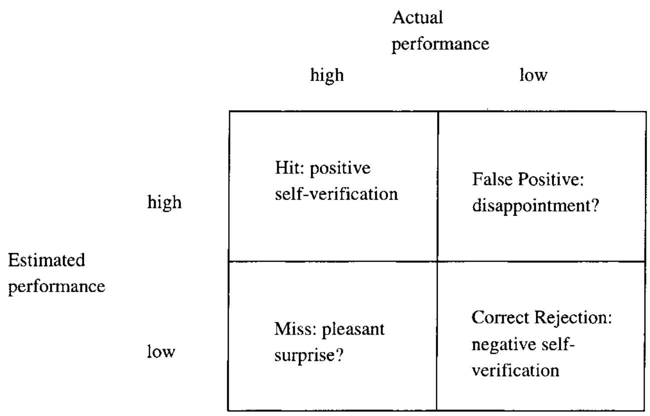


Figure 3. Rational choice between optimism and pessimism.

assume that people seek to maximize hedonic value. If the valued difference between hits and false positives is greater than the difference between misses and correct rejections, optimism is reasonable.

At the level of the individual, neither optimism, pessimism, nor the shift from one to the other violates any generic statistical logic. By weighting the desirability of different kinds of errors, we find that individuals reason well within the Neyman–Pearson theory of statistical decision making (see Hays, 1973, pp. 332–388). Null hypothesis significance testing, as conceived by Fisher (1935) and practiced by most investigators, is a poor way of judging the rationality of social perception (Krueger, 1998a). To require individuals to make estimates identical to the group average of 50% would be to commit the logical fallacy of division. What is true for the group need not be true for the individual group member.

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,* 1621–1630.

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74,* 955–966.

Bobrow, J., Nathan, N., Fisher, S., Covino, W. A., Orton, P. Z., Bobrow, B., & Wever, L. (1989). *Praxis II: NTE core battery: National Teacher Examinations preparation guide.* Lincoln, NE: Cliffs Notes.

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition, 4,* 353–376.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts.* London: Guilford Press.

Clement, R. W., & Krueger, J. (2000). The primacy of self-referent information in perceptions of social consensus. *British Journal of Social Psychology, 39,* 279–299.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

Dunning, D., & Hayes, F. (1997). Evidence for egocentric comparison in social judgment. *Journal of Personality and Social Psychology, 71,* 213–229.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). New York: Macmillan.

Fisher, R. A. (1935). *The design of experiments.* Edinburgh, Scotland: Oliver & Boyd.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland, 15,* 264–263.

Gilovich, T., Griffin, D. W., & Kahneman, D. (2002). *Heuristics of judgment: Extensions and applications.* New York: Cambridge University Press.

Goldberg, L. R. (1991). Human mind versus regression equation: Five contrasts. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 1, pp. 173–184). Minneapolis: University of Minnesota Press.

Haxby, J., Parasuraman, R., Lalonde, F., & Abboud, H. (1993). SuperLab: General-purpose Macintosh software for human experimental psychology and psychological testing. *Behavior Research Methods, Instruments, & Computers, 25,* 400–405.

Hays, W. L. (1973). *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart & Winston.

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195–222). Thousand Oaks, CA: Sage.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). Oxford, England: Oxford University Press.

Klar, Y., & Giladi, E. E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25,* 585–594.

Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79,* 216–247.

Krueger, J. (1998a). The bet on bias: A foregone conclusion? *Psycoloquy, 9*(46). Retrieved January 1, 2001, from http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?9.46

Krueger, J. (1998b). Enhancement bias in the description of self and others. *Personality and Social Psychology Bulletin, 24,* 505–516.

Krueger, J. (1998c). On the perception of social consensus. *Advances in Experimental Social Psychology, 30,* 163–240.

Krueger, J. (2000a). Individual differences and Pearson's *r*: Rationality revealed? *Behavioral and Brain Sciences, 23,* 684–685.

Krueger, J. (2000b). Three ways to get two biases by rejecting one null. *Psycoloquy, 11*(51). Retrieved January 1, 2001, from http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?11.051

Krueger, J. (2001). Null hypothesis significance testing: On the survival of a flawed method. *American Psychologist, 56,* 16–26.

Krueger, J., & Stanke, D. (2001). The role of self-referent and other-referent knowledge in perceptions of group characteristics. *Personality and Social Psychology Bulletin, 27,* 878–888.

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

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,* 1121–1134.

McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York: Wiley.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5,* 241–301.

Nisbett, R., & Ross, L. (1980). *Human inference.* Englewood Cliffs, NJ: Prentice-Hall.

Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin, 79,* 294–309.

Shepperd, J. A. (1993). Student derogation of the Scholastic Aptitude Test:

Biases in perceptions and presentations of College Board scores. *Basic and Applied Social Psychology, 14,* 455–473.

Shepperd, J. A., Ouellette, J. A., & Fernandez, J. K. (1996). Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. *Journal of Personality and Social Psychology, 70,* 844–855.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.

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

Swann, W. B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review, 91,* 456–477.

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

Wedell, D. H., & Parducci, A. (2000). Social comparison: Lessons from basic research on judgment. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 223–252). New York: Kluwer Academic/Plenum.

Appendix

Sample Test Items

Difficult Test

The Greek slave, Aesop, (A) had the ability to (B) translate (C) into memorable stories the idiosyncrasies , (D) faults, and virtues of the people around him. No error. (E)	Confidence that **your own answer** is correct. (1–8)	Confidence that **majority of Brown students** would answer question correctly. (1–8)

Easy Test

The effect of the libraries (A) campaign to encourage children's (B) reading has been (C) overwhelmingly successful according to the fact-finding team. (D) No error. (E)	Confidence that **your own answer** is correct. (1–8)	Confidence that **majority of Brown students** would answer question correctly. (1–8)

Received August 24, 2000
Revision received June 4, 2001
Accepted June 4, 2001 ■