THE JOURNAL OF BEHAVIORAL FINANCE, 9: 187–192, 2008

ISSN: 1542-7560 print / 1542-7579 online
DOI: 10.1080/15427560802539433

Routledge
Taylor & Francis Group

Commentary

Predicting Relative Performance in Economic Competition

Jeffrey Hales
Georgia Institute of Technology

Steven J. Kachelmeier
University of Texas at Austin

Recent research in psychology has challenged the notion of a systematic "better-than-average" bias in predicted relative performance. Extending this research, we show that in an incentive-compatible experimental competition with widespread variation in abilities, those who are better than average tend to overestimate their actual performance, while those who are worse than average tend to *underestimate* their actual performance. We then demonstrate that this symmetric miscalibration at both sides of the relative performance distribution can facilitate optimal choices of costly insurance to mitigate performance-based risks. In contrast to the findings of prior research, we find that, after taking this insurance into account, participants are no worse off when economic risks are based on relative performance than when risks are assigned randomly in a similarly structured control setting.

Keywords: Experimental economics, Prediction, Relative performance, Better-than-average bias

INTRODUCTION

Research in behavioral economics and finance has relaxed the standard assumptions of economic theory by incorporating the behavioral notion that people believe that they are better than average (e.g., Gervais and Odean [2001], Hvide [2002], Simon and Houghton [2003], Hayward et al. [2006]). The tendency for researchers to accept relative overconfidence as a stylized fact is typified by Camerer and Lovallo [1999, p. 306], who conclude that "psychological studies show that *most people are overconfident* about their own relative abilities (emphasis added)." This miscalibration, as they and others show, causes individuals to earn less, on average, when success depends on skill rather than chance.[1]

However, recent advances in social psychology challenge the notion of a general "better-than-average" (BTA) bias. These studies argue that the BTA phenomenon is a subset of a more general tendency for people to confuse perceived absolute performance with perceived relative performance, leading not only to overestimation of relative performance (a BTA effect) for easy tasks but also to *underestimation* of relative performance for difficult tasks (a worse-than-average effect, or WTA). We extend this research by testing the premise that tasks with wide distributions of perceived task difficulty should lead to *simultaneous* BTA and WTA miscalibration by different elements of the population *within the same task*.

THE BTA EFFECT AND RECENT CHALLENGES

While Camerer and Lovallo's [1999] assertion of a general BTA effect accurately summarizes the literature at the time their article was published, social psychology has subsequently challenged the notion of a systematic BTA bias (e.g., Kruger [1999], Klar and Giladi [1999], Klar [2002],

Address correspondence to Jeffrey Hales, Assistant Professor, Georgia Institute of Technology, 800 West Peachtree Street NW, Atlanta, GA 30308. E-mail: jeffrey.hales@mgt.gatech.edu

Moore and Kim [2003]). While the cognitive explanations set forth in these studies differ somewhat, their common theme is that people overgeneralize the extent to which their absolute ability predicts their ability relative to others. Kruger [1999] explains this phenomenon as a form of anchoring (Tversky and Kahneman [1974], Jacowitz and Kahneman [1995]), in which individuals assess relative ability by anchoring on likely absolute ability, adjusting insufficiently for the ability of others. Klar's [2002] "singular-target-focus theory" advances an even simpler premise: When asked to form assessments of any quality relative to others, people tend to drop the "relative to others" and instead make absolute assessments.[2]

To demonstrate the point, the standard approach in these studies is to manipulate task difficulty, showing that too many people believe they are better than average at easy tasks, such as driving or riding a bicycle, and too many people believe they are worse than average at difficult tasks, such as juggling (Kruger [1999]). Interestingly, even Camerer and Lovallo's [1999] results, which many cite as support for a general BTA effect, are logically consistent with the conclusions of these BTA challenges, when one takes into account that Camerer and Lovallo restricted participation to men (see their footnote 7) and found the largest BTA effect in a "self-selection" treatment in which recruiting materials emphasized to potential participants that performance on sports trivia would determine their compensation. Given the well-documented male bias in sport participation and fan interest (e.g., Messner [1988], Wann et al. [1998]), their participants likely believed that a sports trivia contest would be fairly easy.

Indeed, this explanation is consistent with Camerer and Lovallo's [1999] explanation that their participants fail to fully calibrate the extent to which other participants are also good at sports trivia (which they term "reference group neglect"). In fact, they go on to note that one testable implication of reference group neglect is that it "predicts an opposite bias when [people] lose tournaments and consider whether to enter 'consolation' tournaments" (p. 316). Our study explores this implication.

Simultaneous Over-/Underestimation of Relative Performance

Given the prior demonstrations of *systematic* BTA effects for clearly easy tasks and WTA effects for clearly difficult tasks, we test whether BTA and WTA miscalibration can occur *simultaneously* among different elements of the same population when tasks have wide distributions of perceived ability. Sports trivia provides such a task, given that the general population includes avid sports fans who are likely to find such a task "easy," as well as those who do not generally follow sports (and so are likely to find sports trivia to be more foreboding).

To be sure, avid sports fans *should* do better at sports trivia than non-sports fans. Nevertheless, even if those who perceive themselves to be above or below average in a task with widely distributed abilities tend to be correct *relative to the median*, an excessive fixation on perceived absolute ability should lead to greater extremity in such beliefs than is warranted by actual relative performance differences. In other words, if people underweight the presence of others like themselves in the population, the predicted relative performance distribution will be more extreme (i.e., wider at the tails) than the actual relative performance distribution.

Mitigating Risk

In many situations involving social comparisons, parties can take actions to mitigate the risk of poor relative performance. In contemplating a class examination, for example, weaker students should study more than their stronger classmates. Similarly, professionals, such as auditors and physicians, can mitigate various risks by working harder or investing in insurance or other technologies. In each case, the prudence of these costly choices depends to some degree on one's predicted relative performance.[3] To our knowledge, no other study on the BTA or (WTA) effect has specifically examined the implications of allowing participants to "hedge their bets." Given the symmetric nature of the miscalibration we predict (i.e., overly extreme relative performance predictions at both sides of the distribution) along with the insurance options we provide, it is an open question whether misestimating one's relative performance will necessarily impair one's profitability.

EXPERIMENTAL DESIGN

Task

Forty undergraduate student volunteers recruited from university business courses participated in the experiment. In each of 20 rounds, participants evaluated their willingness to accept a risky prospect in which they would incur one of five costs, defined at levels of 10, 25, 50, 75, or 90 points (average of 50). Costs and all other compensated decisions were denominated in "points," with the instructions explaining that we would convert accumulated points to cash compensation after analyzing all responses.[4]

In 10 of the 20 rounds, participants were informed that costs would be assigned based on the relative accuracy of participants' numerical answers to questions involving sports trivia. For instance, an example question in the instructions asked participants to guess the number of victories earned by the university men's baseball team in a particular season. For each question, we ranked the participants' numerical answers in order of accuracy, aggregating across all participants from all experimental sessions we conducted. After calculating performance ranks for each question, we split these ranks

into quintiles and assigned cost levels. For example, answers in the most accurate quintile received the lowest cost, 10 points, and those in the least accurate quintile received the highest cost, 90 points.

To provide a benchmark measure of predicted relative performance, before viewing each skill-based round's trivia question, participants answered the question, "What percent of participants do you think will provide a more accurate answer than you on the question you will answer in this round?"[5] We obtain incentive-compatible measures of the willingness of our participants to bear risks in this task by asking participants to make two decisions tied to their compensation. First, after providing their initial predictions as described above, participants specified a minimum acceptable price (MAP) in each round, which was the lowest amount of points the participant would be willing to accept in exchange for bearing the cost he or she would be assigned that round. Following Becker et al. [1964], participants received a computer-generated price as revenue for any given round if and only if the randomly generated price (determined *ex post*) exceeded (or was equal to) the participant's MAP.

After specifying a MAP, participants chose one of five "cost reduction" options that would apply in the event of a successful MAP. Specifically, participants could pay 0, 9, 18, 27, or 36 points up front (essentially, an insurance premium) to reduce that round's assigned cost by nothing, 20%, 40%, 60%, or 80%, respectively. For example, if a participant were assigned a cost of 90 points (the highest level) but had paid 36 points to reduce costs by 80%, that participant would end up with a net cost of 54 points, calculated as $90 - 0.80(90) + 36 = 54$. Thus, the profit-maximizing option is to pay nothing to reduce costs if a participant's assigned costs are at one of the two lowest levels, or to pay 36 points to reduce

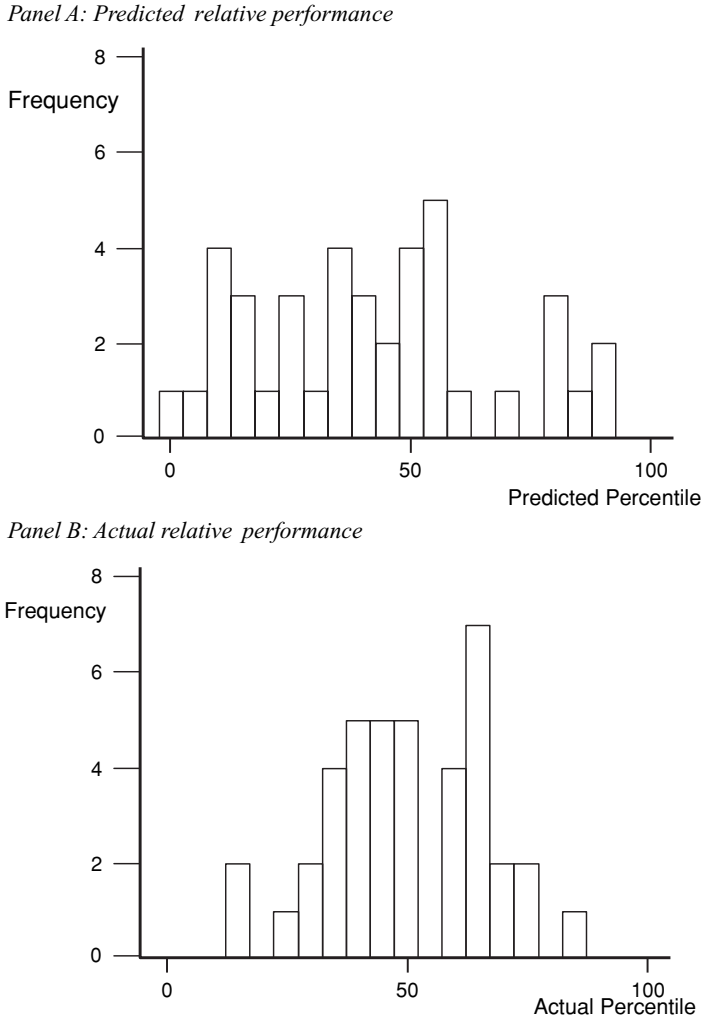

FIGURE 1 Distributions of predicted and actual relative performance. *Note:* Panel A is a frequency histogram of the distribution of predicted relative performance averaged over the 10 skill-based rounds. Panel B contains a similar histogram for the distribution of actual performance in those same rounds. Both graphs use 21 intervals between 0 and 100 (n = 40).

costs by the maximum amount (80%) if assigned costs are at one of the three highest levels.

RESULTS

Predicted Relative Performance

We begin by examining participants' average prediction of their relative performance in the 10 skill-based rounds. In contrast to prior studies showing a general BTA effect, our data reveal an overall WTA effect, if anything, in that the average participant expects to be *out-performed* by 58.8% of other participants. As we explain next, however, a deeper examination of the data reveals a pattern of simultaneous over- and underestimation of relative performance by different elements of the population.

Consistent with our predictions, we find that the direction and magnitude of miscalibrated predicted performance depends on whether participants expect to be above or below average ($F = 22.02$; $p < 0.001$).[6] Participants who predict below-average skill-based performance underestimate their actual performance by an average of 17% ($t = 5.55$; $p < 0.01$). Conversely, participants who predict above-average performance *overestimate* their performance by an average of 7.5% ($t = 1.78$; $p = 0.05$). Thus, while generally able to predict their position relative to the median, both above- and below-average performers are more extreme in their predictions than in their actual performance.

A more systematic way to test for simultaneous BTA and WTA biases is to compare the spreads of the predicted vs. actual performance distributions. By construction, the actual distribution of percentile-ranked relative performance within any given skill-based round is uniform from 0 to 100. When averaged across multiple rounds, the actual distribution of relative performance approximates a typical normal distribution centered at 50. If participants are miscalibrated in the manner we predict, we expect to see greater density in the tails of the predicted performance distribution than in the actual performance distribution, reflecting a tendency to predict relative performance (good or bad) that is more extreme than actual relative performance. This difference is apparent in Figure 1, in that the distribution of predicted performance is wider and flatter than the distribution of actual performance ($t = 3.06$; $p < 0.01$).

Mitigating Risk

Participants with lower self-perceptions of relative performance can mitigate their risks by choosing one of five levels of cost reduction (essentially, insurance), with greater levels of reduction costing more up-front points (essentially, greater insurance premiums). To gauge the wisdom of participants' cost-reduction choices, we compare the outcomes of such choices to those that would have obtained from the choices the same participants made in the random-cost rounds that we used as a control condition, thereby separating the influence of participants' performance predictions from the broader influence of risk preferences. As expected, participants make different cost reduction choices in the skill-based rounds than in the random rounds, depending on whether they expect to perform below or above average. Moreover, these different choices lead participants to incur net costs that are significantly lower than if they had simply extended their random-round cost-reduction choices to the skill-based rounds ($p < 0.10$ for both above-average and below-average participants). Thus, notwithstanding the observed BTA/WTA biases for predicted relative performance, participants appear to benefit from their convictions when investing in costly insurance.

Expected Profit

We next ask whether participants as a whole are better or worse off when risks are skill-based than in the control condition with randomly assigned risks. To test this question, we construct an "expected earnings" variable by computing the expected point outcome associated with a participant's actual performance, MAPs, and cost-reduction choices.[7] Each participant provides two observations – expected earnings averaged across all skill-based rounds and expected earnings averaged across all random rounds.

As shown in Table 1, we find that the influence of skill-based vs. random rounds on expected earnings depends on whether participants predict above-average or below-average performance ($F = 3.99$; $p = 0.05$, two-tailed). Participants predicting below-average performance do not experience significantly different expected earnings in the skill rounds than

TABLE 1
Expected Earnings ANOVA.

Source	df	SS	MS	F-stat	p-value
Between-subjects					
Prediction	1	7.35	7.35	0.07	0.797
Error	36	3928.20	109.12		
Within-subjects					
Skill/Random	1	115.89	115.89	2.19	0.148
Prediction × Skill/Random	1	211.23	211.23	3.99	0.053
Error (Skill/Random)	36	1903.58	52.88		
Skill—random difference by prediction level					
	Mean	Median	Std. Dev.	*t-stat*	*p-value*
Below average	−0.91	−0.20	9.44	−0.48	0.634
Above average	6.12	3.90	11.80	1.87	0.086

Note: The table reports an ANOVA on average expected earnings calculated separately across each participant's skill-based and random rounds (within-subjects factor "Skill/Random"), conditional on whether the participant predicts above-average or below-average performance (between-subjects factor "Prediction"). Expected earnings in any given round is measured as the participant's MAP in that round less the expected cost given the participant's average relative skill level and their cost reduction choice in that round. All p-values are two-tailed.

in the random rounds ($t = –0.48$; $p > 0.50$), while participants predicting above-average performance experience marginally higher expected earnings in the skill rounds than in the random rounds ($t = 1.87$; $p = 0.09$, two-tailed). In sum, whereas Camerer and Lovallo [1999] find that BTA effects cause participants to earn less in performance-based tasks than in tasks based on chance alone, we find that individuals do not always suffer impaired profitability in performance-based tasks. The different conclusions likely reflect the insurance option we incorporate in our task, but we would argue that options of this nature are common in business environments.[8]

CONCLUSION

In this study, we show that some tasks can be simultaneously perceived as easy or difficult by different elements of the population, leading to simultaneous overestimation and underestimation of relative performance by individuals who predict themselves to be above or below average, respectively. As a result, the distribution of predicted relative performance can become more extreme (i.e., wider tails at both ends) than the distribution of actual relative performance. In expanding upon this basic finding, we find that participants generally condition insurance decisions on their predicted performance, and that this makes them better off than if they were to make insurance decisions as if they were facing randomly assigned risks.

Regarding profitability, we find that above-average participants fare better in a skill-based condition than do below-average participants. However, we find no evidence that participants are any worse off overall when profits are based on relative performance than in a control condition with randomly assigned costs. There are two reasons for the difference between this result and the findings of Camerer and Lovallo [1999], who show that overestimating relative performance impairs profitability. First, we examine the entire distribution of likely perceived difficulty for a sports trivia task, in contrast to Camerer and Lovallo's study of male sports fans. Examining a wider distribution of participants likely takes some of the "bite" out of reference group neglect. Second, we provide participants with a mechanism to insure against risks, which increases average payoffs in the skill-based rounds relative to what they would have obtained given their insurance decisions in the random rounds.

In summary, our research shows that, in economic competitions with more widely distributed abilities, those who are above average are likely not as good as they think they are, but those who are below average are likely not as bad as they think either. Given the importance of relative performance predictions in many business contexts, we encourage further research regarding when overestimation and underestimation biases are likely to occur, along with the economic implications of these biases.

ACKNOWLEDGMENTS

We are grateful to Don Moore and workshop participants at the Economic Science Association North American Meeting for helpful comments on a prior version of this paper. We also thank the McCombs School of Business for financial support.

NOTES

1. The word "overconfident" can be used to refer not only to assessments of performance *relative to others*, but also to assessments of performance relative to the correct answer and to assessments of the precision of one's information. Like Camerer and Lovallo [1999], our focus in this paper is on assessments of performance relative to others. To avoid confusion, we use the term "BTA effect" when referring to a belief that one is better than average.
2. Klar's [2002] model applies to a myriad of possible qualities, such as happiness, intelligence, irritation, and content, not just to ability. For example, Klar and Giladi [1999] show that participants interpret the question "How content are you relative to your peers?" in terms of the simpler question "How content are you?"
3. In certain professional settings, performance relative to others is particularly important to the extent that organizational and legal standards use relative performance as a benchmark for determining standards of practice.
4. We used a linear formula that would result in compensation ranging from $5 and $50 with an average of $25.
5. In the other 10 rounds within each experimental sequence, participants were informed that the same five cost levels would be assigned randomly. The random rounds serve as an experimental control, enabling us to better control for other effects (such as risk preferences or gender). To minimize order effects, the rounds alternated between the random and skill-based cost assignments. In the random rounds, participants answered the question, "What percent of participants do you think will draw a number lower (i.e., luckier) than you in this round?"
6. All tests are one-tailed, conditional on predictions, unless otherwise specified.
7. Expected earnings differ from actual earnings because the latter also involves a stochastic element (luck) due to the random price draws necessary to implement the Becker et al. [1964] technique. Thus, in the skill-based

rounds, the "expected earnings" metric captures what a participant could expect to earn on average, given actual choices and performance. In the random rounds, expected earnings are calculated based on expected values of costs and cost reductions, ignoring the actual and uncontrollable random realizations of these values.

8. We also tested for gender-based biases. The inclusion of men and women in this study provides a broad distribution of perceived task difficulty for a sports-trivia contest, but also raises the possibility that our results may reflect a more general "gender effect," as opposed to the performance-based explanations that motivate our predictions. However, when we test explicitly for the effect of gender on the degree of miscalibration in relative performance predictions, the gender effect is negligible ($F = 0.26$; $p > 0.50$). Thus, our results appear to be driven by a general tendency for participants to hold beliefs about their relative ability that are too extreme, rather than a more basic effect of participant gender.

REFERENCES

Becker, G. M., M. H. DeGroot and J. Marschak. "Measuring Utility by a Single-response Sequential Method." *Behavioral Science,* 9, (1964), pp. 226–232.

Camerer, C. and D. Lovallo. "Overconfidence and Excess Entry: An Experimental Approach." *The American Economic Review,* 89, (1999), pp. 306–318.

Gervais, S. and T. Odean. "Learning to Be Overconfident." *Review of Financial Studies*, 14, (2001), pp. 1–27.

Hayward, M. L. A., D. A. Shepherd and D. Griffin. "A Hubris Theory of Entrepreneurship." *Management Science*, 52, (2006), pp. 160–172.

Hvide, H. K. "Pragmatic Beliefs and Overconfidence." *Journal of Economic Behavior & Organization*, 48, (2002), pp. 15–28.

Jacowitz, K. E. and D. Kahneman. "Measures of Anchoring in Estimation Tasks." *Personality and Social Psychology Bulletin*, 21(11), (1995), pp. 1161–1166.

Klar, Y. "Way Beyond Compare: Non-selective Superiority and Inferiority Biases in Judging Randomly Assigned Group Members Relative to Their Peers." *Journal of Experimental Social Psychology*, 38(4), (2002), pp. 331–351.

Klar, Y. and E. E. Giladi. "Are Most People Happier Than Their Peers, or Are They Just Happy?" *Personality and Social Psychology Bulletin*, 25(5), (1999), pp. 585–594.

Kruger, J. "Lake Wobegon Be Gone! The "Below-Average Effect" and the Egocentric Nature of Comparative Ability Judgments." *Journal of Personality and Social Psychology*, 77, (1999), pp. 221–232.

Messner, M. A. "Sports and Male Domination: The Female Athlete as Contested Ideological Terrain." *Sociology of Sport Journal,* 5, (1988), pp. 197–213.

Moore, D. A. and T. G. Kim. "Myopic Social Prediction and the Solo Comparison Effect. *Journal of Personality and Social Psychology*, 85(6), (2003), pp. 1121–1135.

Simon, M. and S. M. Houghton. "The Relationship between Overconfidence and the Introduction of Risky Products: Evidence from a Field Study." *Academy of Management Journal*, 46(2), (2003), pp. 139–149.

Tversky, A. and D. Kahneman. "Judgment under Uncertainty: Heuristics and Biases." *Science,* 185, (1974), pp. 1124–1131.

Wann, D. L., M. P. Schrader, J. A. Allison and K. K. McGeorge. "The Inequitable Newspaper Coverage of Men's and Women's Athletics at Small, Medium, and Large Universities." *Journal of Sport and Social Issues*, 22, (1998), pp. 79–87.