Journal of Personality and Social Psychology
2008, Vol. 95, No. 2, 253–273


0022-3514/08/$12.00 DOI: 10.1037/0022-3514.95.2.253

# Are People Excessive or Judicious in Their Egocentrism? A Modeling Approach to Understanding Bias and Accuracy in People's Optimism

Paul D. Windschitl, Jason P. Rose, Michael T. Stalkfleet, and Andrew R. Smith
University of Iowa

People are often egocentric when judging their likelihood of success in competitions, leading to overoptimism about winning when circumstances are generally easy and to overpessimism when the circumstances are difficult. Yet, egocentrism might be grounded in a rational tendency to favor highly reliable information (about the self) more so than less reliable information (about others). A general theory of probability called extended support theory was used to conceptualize and assess the role of egocentrism and its consequences for the accuracy of people's optimism in 3 competitions (Studies 1–3, respectively). Also, instructions were manipulated to test whether people who were urged to avoid egocentrism would show improved or worsened accuracy in their likelihood judgments. Egocentrism was found to have a potentially helpful effect on one form of accuracy, but people generally showed too much egocentrism. Debias instructions improved one form of accuracy but had no impact on another. The advantages of using the EST framework for studying optimism and other types of judgments (e.g., comparative ability judgments) are discussed.

*Keywords:* egocentrism, optimism, likelihood judgment, comparative judgment, shared-circumstance effect

Being able to judge—with some degree of accuracy—how your skills stack up against your peers or how good your chances are of winning in a competitive context is critical. If you underestimate your relative skills and develop an overly pessimistic outlook, you may forego wonderful opportunities because of a fear of failure, or you may waste time and resources overpreparing for competitive tasks for which you are already adequately prepared. If you overestimate your skill and develop rosy expectations for success, you may pick unwise challenges or may fail to adequately prepare for a task and may eventually find yourself suffering a potentially avoidable defeat.

Recent research suggests that when people are judging their comparative ability or estimating the likelihood of winning in a competitive task, they tend to be overly positive when the task is generally easy and overly negative when the task is generally hard (e.g., Burson, Larrick, & Klayman, 2006; Endo, 2007; Larrick, Burson, & Soll, 2007; Kruger, 1999; Moore & Kim, 2003; Moore & Small, 2007; Rose & Windschitl, 2008; Windschitl, Kruger, & Simms, 2003). One reason for this is *egocentrism,* which in the present context can be defined as a tendency for thoughts about the self and about self-relevant information to carry more weight in shaping comparative or likelihood judgments than do thoughts about others and other-relevant information (see Chambers & Windschitl, 2004). Indeed, Kruger (1999) found that when people estimate how good they are at a task relative to other people, they tend to think egocentrically; that is, they consider how skilled (or unskilled) they are at the task more so than how skilled (or unskilled) other people are at the task. Therefore, when a task is easy (e.g., operating a computer mouse), people judge their ability at the task to be better than average. However, when the task is difficult (e.g., sculpting human figures from clay), people judge their ability at the task to be worse than average. Windschitl et al. (2003) illustrated that this egocentrism also influences optimism (likelihood judgments) about the outcomes of competitions (see also Kruger, Windschitl, Burrus, Fessel, & Chambers, 2008; Moore & Cain, 2007; Moore & Kim, 2003; Moore & Small, 2007; Rose & Windschitl, 2008). When participants believed they would play a fellow participant in a trivia competition, they tended to be overly optimistic if the relevant trivia category was easy (e.g., current events) and overly pessimistic if the category was difficult (e.g., baroque music). This tendency to be overoptimistic when shared circumstances are easy and overpessimistic when they are difficult is called the *shared-circumstance effect* (SCE; Windschitl et al., 2003). In support of the idea that egocentrism is involved in these SCEs, path analyses showed that participants' likelihood judgments about winning were primarily a function of how much knowledge they thought they had about a category and not a function of how much knowledge they thought their competitor had.

At first blush, this egocentrism seems problematic. Indeed, something must be amiss when both competitors in a pair are highly confident that they will beat each other on the easy trivia categories and will lose to each other on hard categories. Yet, there are different versions of egocentrism, and some versions are more defensible (perhaps rational) than others (Chambers & Windschitl, 2004; Windschitl et al., 2003). One type of an egocentrism account

Paul D. Windschitl, Jason P. Rose, Michael T. Stalkfleet, and Andrew R. Smith, Department of Psychology, University of Iowa.

We thank Gregg Oden for his helpful advice regarding this research. This work was supported by Grants SES 03-19243 and SES 07-20349 to Paul D. Windschitl from the National Science Foundation.

Correspondence concerning this article should be addressed to Paul D. Windschitl, Department of Psychology, University of Iowa, Iowa City, IA 52242. E-mail: paul-windschitl@uiowa.edu

characterizes egocentrism as a chronic attention bias—people are prone to attend to self-relevant information more than they attend to information about others, even when the latter information is equally valuable for making a good judgment. This type of egocentrism does not appear to have an immediate, rational basis. However, another type of egocentrism account assumes that people tend to have more knowledge about themselves than about others and that any assessments they make of themselves would tend to be more reliable than assessments they make of others. As a general statistical principle, reliably measured variables tend to be better predictors of outcome variables than are unreliably measured variables. Likewise, it seems warranted to expect that more reliable self-estimates would be better predictors of true comparative standing or competition outcomes than would less reliable estimates of others. Therefore, when people need to make a comparative judgment, they have some justification for giving more weight to assessments about the self than to assessments about others (see Burson & Klayman, 2006; Chambers & Windschitl, 2004; Kruger et al., 2008; Moore & Cain, 2007; Moore & Small, 2007; Ross & Sicoly, 1979).

Recent research by Kruger et al. (2008) demonstrated that when people are asked to predict which of two competitors will win a competition, people's weighting of evidence regarding the competitors differs as a function of how much information they have about the competitor. For example, when participants imagined a high school acquaintance whom they knew quite well competing against an unfamiliar student in a trivia competition, participants predicted that the high school acquaintance would win the easy categories and would lose hard ones. Path analyses confirmed that participants primarily based their predictions on their judgments of how much their high school acquaintance knew about the categories and not on their judgments of how much the unfamiliar student knew. Hence, not only is there a reasonable rationale for people to weight self-assessments more than other-assessments because of a difference in the amount of the information underlying the assessments (and therefore the reliability of those assessments) but also research lends support to the idea that assessments based on more rather than on less information do carry more weight in people's comparative judgments–consistent with rational discounting.

Yet, although the research by Kruger et al. (2008) is consistent with the notion that the weights people give to self-assessments and other-assessments are at least somewhat sensitive to information quality, the research did not address whether people give an optimal set of weights to self-assessments versus other-assessments, nor did the research address how the weights should shift under different conditions. Fully addressing these issues requires that researchers explicitly measure competition outcomes and judgment accuracy, which Kruger et al. (2008) did not do. Addressing these issues also requires a model that formalizes how self-assessments and other-assessments might be used in generating likelihood judgments. Although egocentrism has been discussed in numerous articles regarding optimism, comparative optimism, and comparative judgment, we knew of no formalized models of egocentrism regarding these types of judgments (see e.g., Alicke & Govorun, 2005; Blanton, Axsom, McClive, & Price, 2001; Chambers & Windschitl, 2004; Endo, 2008; Kruger, 1999; Moore & Small, 2007; Pahl & Eiser, 2007; Ross & Sicoly, 1979; Weinstein & Lachendro, 1982).

We had four main goals for this project. The first goal involved bridging research on egocentrism in judging optimism with research concerning generic likelihood judgment. More specifically, we sought to apply an existing likelihood-judgment model to provide a formalized representation of how self-assessments and other-assessments are used when people estimate their likelihood of winning in a competitive situation. The second goal was to use this model and new judgment data to estimate the extent to which people tend to weight self-assessments and other-assessments in three competitions (one involving a trivia quiz and two involving a listing task). The third goal was to use the model, the new judgment data, and the new outcome data to estimate the extent to which people should have weighted self- and other-assessments in the competitions. Hence, whereas the second goal involved a descriptive analysis of judgments, the third goal involved a prescriptive analysis (i.e., how people could optimize the accuracy of their likelihood judgments). Finally, the fourth goal was to examine, within the context of the model and the three competitions, how attempts to debias egocentrism influence judgment processes and the resulting accuracy of people's likelihood judgments.

## Goal 1: A Formalized Model of Egocentrism in Optimism

As stated above, one main goal of this project was to apply an existing likelihood judgment model to provide a formalized representation of how self-assessments and other-assessments are used when people gauge their optimism about winning. We note that judging one's likelihood of succeeding in a competitive context is simply a special case of a generic likelihood judgment. Significant modeling work has already been conducted regarding people's generic likelihood judgments. Perhaps most prominent is the work on support theory (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994; see also Brenner, Koehler, & Rottenstreich, 2002). Our modeling of egocentrism involves an extension of support theory that was developed by Idson, Krantz, Osherson, and Bonini (2001). To simplify matters, when we discuss Idson et al.'s model, we focus only on binary cases (as opposed to polychotomous cases) in which there are only two possible outcomes or hypotheses (e.g., there are two competitors in a competition, exactly one of whom will win).

Idson et al.'s (2001) extension of support theory was developed to better account for binary noncomplementarity (or nonadditivity). Previous research on binary noncomplementarity has illustrated that when people are aware that exactly one of two hypotheses (A and B) are correct, and when both hypotheses have low evidential support, the sum of the mean estimate from people asked to judge the probability of A and the mean estimate from people asked to judge the probability of B will often systematically fall below 1.0 (Macchi, Osherson, & Krantz, 1999; see also Fox & Levav, 2000; McKenzie, 1998; Windschitl, 2000; Yamagishi, 2002). For example, if people know that two terrible basketball teams will be playing against each other, the sum of the average probabilities regarding Team A and Team B winning may fall below 1.0. However, when the evidential support for the two hypotheses is strong, the relevant sum might systematically exceed 1.0. These instances of noncomplementarity (and closely related findings, e.g., see Eiser, Pahl, & Prins, 2001; Klar & Giladi, 1997; Lehman, Krosnick, West, & Li, 1992; Moore & Kim, 2003; Windschitl, Conybeare, & Krizan, 2008; Windschitl et al., 2003) can be generally attributed to *focalism:* a tendency to base a probability judgment primarily on the evidence directly relevant to the focal hypothesis (i.e., the hypothesis for which a probability estimate is sought) rather than the alternative hypothesis. There are actually several versions of focalism accounts (see Chambers & Windschitl, 2004), but each one ultimately stipulates that probability judgments will reflect assessments of evidence for the focal hypothesis more so than assessments of evidence for the alternative hypothesis. Idson et al.'s model was designed to represent the same pattern in a formal (mathematical) fashion. In doing so, it offers some insights, convenience, and precision beyond the verbally articulated focalism accounts. In this article, we apply Idson et al.'s model for representing egocentrism as well as focalism.

A key point behind Idson et al.'s (2001) model is that if a judge reflects primarily on the evidence regarding the focal hypothesis (and neglects the evidence regarding the alternative hypotheses), the evidence regarding the focal hypothesis must be compared with something in order for the evidence to seem strong or weak. In their mathematical model, Idson et al. referred to the something as the constant K. Idson et al. (2001, p. 229) stated,

> Possibly, K may be interpreted as a default value of contrary evidence strength, subject to influence by context and frame, or it may simply be a normalization constant used to convert the open-ended evidence scale into a probability scale.

Consider a case in which a judge is given some pregame statistics for two terrible basketball teams and is asked to estimate the likelihood that Team A will win. If the judge exhibits heavy focalism, he or she would compare the Team A statistics with some internal or default standard for basketball statistics (K), would thereby judge the Team A statistics to be weak, and would then make this the primary basis for his or her probability judgment. Additional perspectives as to what constitutes or shapes K can be gleaned from other literatures. For example, a perspective shaped by norm theory might suggest that K is a norm computed from representations that are evoked by a particular stimulus and by category exemplars or preexisting frames of reference (Kahneman & Miller; 1986). Also relevant would be the work of Giladi and Klar (2002). They posited that when people make direct comparison judgments, they often compare items with a very general standard that is based on stored exemplars from the item's category—even when doing so is not appropriate. We discuss the Giladi and Klar approach and its similarity to Idson et al.'s model later in the article. For the remainder of the article, we adopt a conceptualization of K as a default value or general internal standard.

We henceforth refer to Idson et al.'s (2001) model and our application of it as extended support theory (EST), as coined by Idson et al. Rather than describe their model in its original terminology, we slightly adapt the terminology and use notations more suitable for the present article. The EST model assumes that the estimated probability of a focal hypothesis in a binary case (call it *A*, with the alternative being *B*) is a function of two things: (a) how the perceived evidence or support for *A* compares with the support for *B* and (b) how the support for *A* compares with K. The former comparison is represented as

$$\frac{s(A)}{s(A) + s(B)}, \tag{1}$$

with $s$ ($A$) read as *support for A*. The latter comparison, in which support for *A* is compared with K (the default or general standard), is represented as

$$\frac{s(A)}{s(A) + K}. \tag{2}$$

A weighting parameter (λ; lambda) determines weight for the former versus the latter comparison in shaping the probability judgment. Hence, the full model for the judged probability of *A* rather than *B* is then represented as

$$\mathrm{P}(A,B) = \lambda\frac{s(A)}{s(A) + s(B)} + (1 - \lambda)\frac{s(A)}{s(A) + K}. \tag{3}$$

From a prescriptive perspective, it would seem that λ should be 1.0. That is, people should base their probability judgment on how the support for the focal hypothesis (*A*) compares with the support for the alternative (*B*), not whether the support for the focal hypothesis is high or low in general terms (i.e., high or low relative to the default, K). However, consistent with the above discussion, research on focalism suggests that from a descriptive perspective, λ is often less than 1.0 and that people's probability judgments are in part determined by how the support for the focal hypothesis compares with some general default. Continuing with the basketball team example, when judging the probability that a terrible basketball team will beat an equally terrible team, people would often estimate lower than .5 because the support for the focal basketball team is low in general (i.e., relative to the general or default standard).

Now that we have explained how EST can be useful for representing focalism, it is not difficult to extend it to egocentrism (see Brenner et al., 2002). Assume a binary case in which a person is estimating the probability that he or she will win a competition against another individual. With this case in mind, egocentrism would—from a traditional perspective—be defined as a tendency for self-assessments (about strength or support or performance) to have greater weight than do other-assessments in predicting people's probability judgments. However, previous discussions of egocentrism have not established a formalized representation of this egocentrism. Our application of EST formalizes how self- and other-assessments might relate to probability judgment in an egocentric fashion and how such patterns would relate to SCEs. Namely, for people exhibiting egocentrism (or focalism, or both), λ is less than 1.0, and their optimism about winning is partly a function of whether their self-assessed strength, support, or performance is generally high or low relative to a global or default standard, K. If the self-assessment is higher than the global standard (e.g., because all participants in a competition estimate their performance on an easy category to be generally high), these people would tend to be overoptimistic. If the self-assessment is lower than the global standard, these people would tend to be overpessimistic.

It is important to reiterate that various egocentrism accounts have been proposed, and each describes a different set of processes that result in egocentric weighting (see Chambers & Windschitl, 2004). As examples, one account characterizes egocentric weighting as the result of an attention bias; another, as a result of differential confidence about self-relevant and other-relevant information; and another, as a result of egocentric anchoring and insufficient adjustment. EST does not directly sort among these egocentrism subaccounts. Yet, EST provides various conceptual insights into egocentrism, and as discussed in the next two sections, EST provides a useful analytic framework for examining the descriptive and prescriptive roles of egocentrism (and focalism) across various contexts.

## Goal 2: The Descriptive Role of Egocentrism

The second goal for this project was to use the EST model and the judgment data to estimate—descriptively—the extent to which people tend to weight self- and other-assessments in three competitions. The three competitions (Studies 1, 2, and 3, respectively) are similar in structure. In all studies, each participant was pitted against a fellow participant. In Study 1, the participants individually completed a trivia quiz for each of eight trivia categories that ranged from generally hard to easy. In Studies 2 and 3, participants completed a time-limited listing task for each of 14 categories that also ranged from generally hard to easy. All participants provided estimates of the likelihood that they would beat their competitor on each category, as well as estimates of the number of items they and their competitor answered or listed correctly for each category (i.e., score estimates). An instructional manipulation (standard or debias) that was used in the experiments is described later.

For each participant in the studies, we determined his or her descriptive level of egocentrism.[1] Specifically, using EST (Equation 3 from above), we solved for $\lambda$. Self- and other-score estimates were used for $s(A)$ and $s(B)$, respectively (see discussion by Koehler, Brenner, & Tversky, 1997, p. 296). Consistent with conclusions from previous work (e.g., Rose & Windschitl, 2008; Windschitl et al., 2003), we expected there to be robust egocentrism. Thus, we expected values of $\lambda$ to be substantially below 1.0. If the value of $\lambda$ estimated from a person's data was 1.0, this would mean that the person's probability estimates were strictly a function of how the score predictions about the self differed from the score predictions about the other person (i.e., no egocentrism). However, if $\lambda$ was 0, this would mean that the person's probability estimates were strictly a function of how the score predictions about the self differed from some default or general expectation (i.e., pure egocentrism). Again, we expected there to be robust egocentrism, with $\lambda$s substantially below 1.0. This egocentrism should also lead to overoptimism about easy tasks in the competition but to overpessimism about hard tasks in the competition, which would yield an SCE.

## Goal 3: The Prescriptive Role of Egocentrism

Our third goal was to use the model, the judgment data, and the outcome data to estimate—prescriptively—the extent to which people should have weighted self- and other-assessments. In other words, we investigated the optimal level of egocentrism for achieving maximal accuracy in likelihood judgments about winning. Using Equation 3 and using a given participant's self and other score estimates, we solved for the value of $\lambda$ that produced likelihood estimates that were maximally correlated with actual wins and loses. Henceforth, we often use the term *prescriptive* $\lambda$ to refer to this optimal value of $\lambda$, whereas we use the term *descriptive* $\lambda$ when referring to the value of $\lambda$ discussed in the above paragraph.

Because the actual outcomes of the competitions are necessarily determined by both the scores from the self and the scores from the other person, it might seem that 1.0 would be the prescriptive value of $\lambda$, reflecting no egocentrism. However, consistent with our earlier discussion, if other score estimates are much noisier and less reliable than are self-score estimates then perhaps a person should discount the importance of their prediction about the other person's score. In other words, perhaps prescriptive $\lambda$ would be substantially less than 1.0.[2]

Continuing with this logic, the optimal or prescriptive weighting for score predictions about the self and the other should vary across competitions. In some types of competitions, a person might have very limited insight into their competitor's performance (e.g., when their competitor is a stranger and the task is novel). Their only method of generating a score prediction for the other person might be through projection (see e.g., Dawes & Mulford, 1996; Hoch, 1987; see also Karniol, 2003; Krueger, 2003; Krueger & Clement, 1994; Ross, Greene, & House, 1977; Sherman, Presson, & Chassin, 1984). Although this method might be sensible, it is not insightful as to how the competitor will score relative to the self. In other types of competitions, a person might have much better insight into their competitor's performance (e.g., when they know the competitor and something about the competitor's suitability for the given task). In this latter case (but not the former), it may be optimal for score predictions about the other person to have as much influence as score predictions about the self ($\lambda$ should be near 1.0).

We assumed that the participants in Study 1 (who were strangers) would be unlikely to have much insight into their competitor's unique strengths and weaknesses at the trivia categories in the competition. For reasons described later, we expected participants' knowledge of their competitors to be better in Study 2 and best in

[1] Because we were examining people's likelihood judgments about themselves winning a competition, differential weighting that was detected was consistent with both egocentrism and focalism. That is, egocentrism and focalism are indistinguishable when the self is part of the focal outcome. Previous research that has manipulated whether the self is part of the focal outcome has shown that egocentrism—apart from focalism—is a substantial contributor to SCEs (Windschitl et al., 2003). For convenience, then, we use the term egocentrism when referring to data patterns reflecting more weight for self-assessments than other-assessments, even though focalism might contribute to such patterns.

[2] The argument that people may be rational in discounting (or giving less weight to) their estimates of the competitor's score presumes that people have not already substantially regressed their estimate of their competitor's score. That is, if participants tend to report regressive estimates about their competitor (i.e., estimates that do not differ much from a baseline), this would suggest their score estimates might already reflect the relative unreliability of information about others. Hence, differential weighting of score estimates about the self and competitor would not be necessary. However, as we report later, although there is some evidence that participants' score estimates regarding others were more regressive than their score estimates about the self, the magnitude of this effect was not sufficient to rule out the potential rational benefits of differential weighting. In short, a more complete statement of the rationality argument would be as follows: Given that people do not always fully regress their scores estimates for others, it may be rational for people to differentially weight self and other scores when making likelihood judgments.

Study 3. Hence, we expected that the optimal value of λ would be lowest in Study 1 and highest in Study 3.

Also, we can determine whether people were more or less egocentric than they should have been, based on a comparison of the prescriptive and descriptive λs. Although we expected that people should prescriptively be somewhat egocentric in making their likelihood judgments, we also expected that their descriptive level of egocentrism would actually exceed what would be prescriptively warranted. This is because egocentrism is a weighting bias that is multiply determined. Although rational discounting might underlie some of the egocentrism of our participants, we also suspected that nonrational causes, such as an egocentric and/or focalistic attentional bias, would augment the differential weighting they exhibited.

Before moving on to Goal 4, we should note that the issue of whether egocentrism is warranted for achieving accuracy in likelihood judgments is somewhat parallel to the issue that has previously been addressed regarding false consensus—namely, the possibility that projection is helpful rather than problematic when a person is estimating the responses or characteristics of others (Davis, Hoch, & Ragsdale, 1986; Dawes & Mulford, 1996; Hoch, 1987, 1988; Krueger & Clement, 1994; Krueger & Zeiger, 1993). We more fully discuss the somewhat parallel nature of these issues in the General Discussion section, but it is important to emphasize here that the issues are distinct from each other. Whereas Davis et al., (1986), Hoch (1987, 1988), Dawes and Milford (1996), and Krueger (Krueger & Clement, 1994; Krueger & Zeiger, 1993) essentially focused on whether a person should use self-estimates as a basis for estimating others, our research focuses on how a person should use self-estimates and other-estimates to predict competition outcomes.

## Goal 4: Do Debias Instructions Help, Hurt, or Have No Impact on Accuracy?

If, as we have suggested, people have rational reasons for some egocentrism but tend to overdo the egocentrism for nonrational reasons (e.g., an attention bias), could people benefit from some debias instructions? That is, if people were reminded to consider both the self and the other score predictions when making a likelihood judgment, would this "clean out" the irrational elements of egocentrism, leaving an optimal (or at least more accurate) set of judgments?

Here it is important to distinguish between two types of accuracy or inaccuracy (see e.g., Epley & Dunning, 2006; González-Vallejo & Bonham, 2007; Yates, 1990, 1994). First, there is mean-level inaccuracy, which is related to the issue of calibration in the overconfidence literature. The SCE is an illustration of mean-level inaccuracy in the sense that people are generally overoptimistic about easy categories and underoptimistic about hard ones. A second type of accuracy or inaccuracy concerns whether people tend to actually win the tasks for which they have given high probability estimates and lose the tasks for which they have given low probability estimates (related to discrimination accuracy; measured in this article by a correlation between likelihood estimates and actual win–loss outcomes). A critical point is that these two types of accuracy are not necessarily influenced by the same factors—they are semi-independent.

We expected that debias instructions would always have a positive influence on mean-level accuracy or calibration. More specifically, we expected that participants who read debias instructions would show significantly reduced SCEs. This is because egocentrism of any type—regardless of whether it is rational or irrational—will tend to produce SCEs. However, we did not expect that debias instructions would necessarily have a positive influence on discrimination accuracy (the within-participant correlations between optimism and win–loss outcomes across categories). For this to occur, people would need to remove just the right amount of egocentrism. That is, they would need to remove the egocentrism due to nonrational causes but not the egocentrism (or its amount) that is rationally justified. Given the many hurdles that there are to decontaminating a bias (see e.g., Anderson, Lepper, & Ross, 1980; Larrick, 2004; Wegener & Petty, 1997; Weinstein & Klein, 1995; Wilson & Brekke, 1994), we expected that people would be rather coarse in adjusting their egocentrism in response to debias instructions. Hence, we expected that even if or when debiasing improved mean-level accuracy, it would not improve discrimination accuracy.

## Summary of Goals

Briefly summarized, our goals for the article were to (a) apply a generic likelihood judgment model in order to provide a formalized representation of how self-assessments and other-assessments influence optimism in competitions, (b) use the model and judgment data to investigate descriptive weighting, (c) use the model, judgment data, and outcome data to investigate prescriptive weighting, and (d) examine how debias instructions influence various forms of accuracy.

## Overview of the Competitions: Studies 1–3

We pursued these goals within three competitive environments: Studies 1, 2, and 3. Whereas Study 1 involved trivia quizzes, Studies 2 and 3 involved listing tasks. All the studies and their competitions were similar in structure and involved the same types of variables. This similarity in structure and variables allows us to make some informative comparisons across studies. The methodology and results for the three studies are described together when possible.

All studies involved one-on-one competitions in which participants individually completed competitive tasks involving categories that ranged from hard to easy. At this point, therefore, participants had direct experience with the tasks but did not know how their competitor had done (nor did they know their own precise scores). After performing the tasks, participants provided likelihood judgments about winning each category, and they estimated the self and other scores for each category (i.e., score estimates). The studies also contained an instruction manipulation (standard or debias) that varied on a between-subjects basis.

A key difference among the studies was the extent to which participants had a valid sense of their competitive advantage or disadvantage for each category. For the trivia categories in Study 1, participants did not have a good sense of their competitive advantage or disadvantage. For Studies 2 and 3, which were run concurrently, we chose a listing task involving cate-

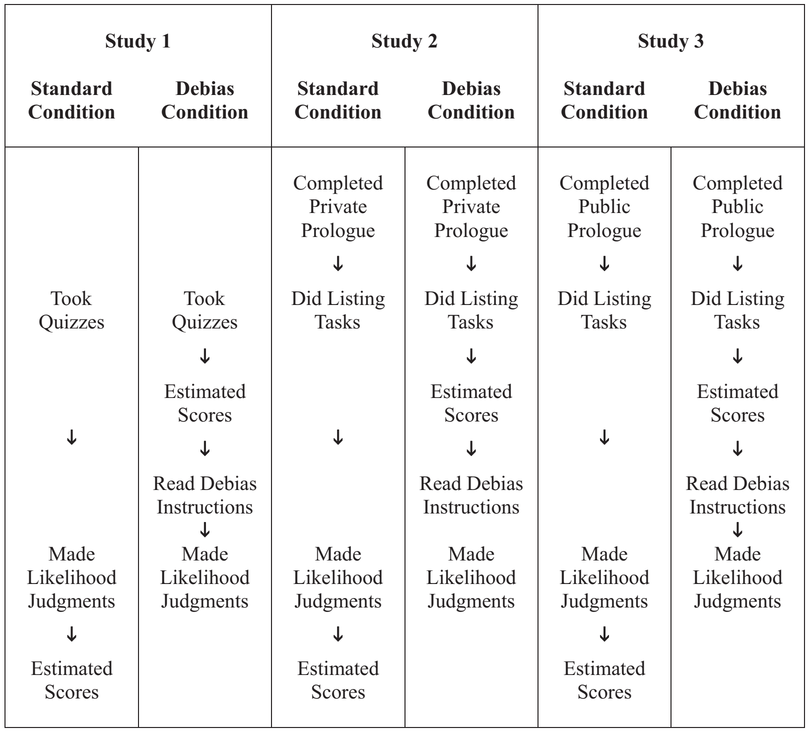


*Figure 1.* A schematic for the ordering of the main procedural elements of Studies 1–3.

gories for which participant would naturally have a better sense of their competitive advantage or disadvantage. In fact, in Study 3, an activity prior to the competition gave participants additional insight about their competitive advantage or disadvantage for each category.

Before describing specific procedures for the studies, a word is in order regarding our debiasing instructions, which may seem heavy handed to the reader. Indeed, we intended for our debias instructions and procedures to be strong, if not heavy handed. Our interest was not in whether modest instructions could effectively debias SCEs. Instead, our reason for including a debias condition was to see how participants' active attempts to avoid egocentrism (prompted by our instructions and procedures) would influence various forms of accuracy. Metaphorically, we led our horse to water because we wanted to learn what would happen at the water.

## Method for Studies 1–3

### *Participants for Studies 1–3*

The participants (*N*s = 56, 58, and 60 for Studies 1, 2, and 3, respectively) were University of Iowa students fulfilling a research exposure component of their elementary psychology course.[3]

### *Procedure and Measures for Study 1*

Figure 1 contains a schematic for the ordering of the key procedural elements of Study 1 (as well as Studies 2 and 3). Two participants (strangers) arrived for each session. Upon arrival, they learned that they were about to compete in a trivia contest against their coparticipant and that depending on their performance, a small amount of money could be earned. Immediately after this, participants individually answered seven multiple-choice questions and one tiebreaker question in each of eight trivia categories. Four of the quiz categories were designed to seem hard for our participants (e.g., South American geography, world's rivers) and four were designed to seem easy (e.g., pop culture, fast food chains; see Windschitl et al., 2003). The subsequent set of procedures differed between standard and debias conditions.

*Standard condition for Study 1.* Participants in the standard condition estimated the likelihood that they would beat their competitor in each of the eight categories ("Please indicate what you think your chances are of winning each category. Please give a numeric likelihood estimate between 0 and 100%."). Instructions made it clear that there would be only one winner per category, with tiebreakers used when needed. Participants then provided score estimates for themselves and their competitor for each category (e.g., "I answered _____ [write in 0–7] items correctly for the Pop Culture category"). The order in which they made the eight

[3] In some analyses reported in this article, the actual *n* is 1 or 2 participants smaller than reported here. This occurred when, for example, a participant failed to respond to a relevant question or provided estimates of zero for both the self and the competitor for multiple categories.

score estimates for the self and the eight estimates for the competitor was counterbalanced.[4]

*Debias condition for Study 1.* To encourage participants in the debias condition to see that both the score estimates about themselves and their competitor were important for their optimism about winning, we had these participants answer the score-estimate questions before providing likelihood judgments. Also, just prior to making the likelihood judgments, participants read very strongly worded debias instructions—telling them that most participants tend to neglect the strengths and weaknesses of their competitor when gauging optimism, and urging them to avoid this tendency (see Appendix).

### *Procedure and Measures for Studies 2 and 3*

Studies 2 and 3 were run concurrently and could be considered two between-subjects conditions from the same study. However, for clarity in exposition, we refer to them as Study 2 and Study 3. For these studies, we constructed listing tasks involving categories for which participants would have a better sense (relative to the categories in Study 1) of whether they were at a competitive advantage or disadvantage compared with their coparticipant (a stranger). Our intuition about our selected categories was verified by informal pilot testing and by results reported later in this article.

The only difference between the method for Study 2 and the method for Study 3 involves a prologue to the actual competition. In the prologue, participants answered two questions about each of 14 categories (the same categories that would later be critical for the actual competition). First, they indicated how good they would be at listing items from a specified category (e.g., the planets in our solar system) within a "short period of time" (1 = *not very good*, 5 = *very good*). Second, they provided a brief reason why they believed they would be good or not good at listing such items. The only difference between Study 2 and Study 3 was that in Study 2, participants answered these prologue questions privately and on paper; in Study 3, participants answered these questions publicly and orally, such that their coparticipant heard each answer. Hence, we intended these publicly stated answers to provide participants in Study 3 with additional insight about their competitive advantage or disadvantage in the actual competition that followed.

After the prologue, all participants learned that they were about to compete in a series of listing contests against their coparticipant and that depending on their performance, a small amount of money could be earned. For these listing contests, we selected 14 categories (from a pretested pool) that ranged from generally difficult (e.g., tool brands) to generally easy (e.g., planets). The order in which participants encountered the categories in the listing task was random. For each of 14 categories, participants were given 30 s to individually list (on paper) as many items from that category as possible. The subsequent set of procedures differed between the standard condition and the debias condition (see Figure 1).

*Standard condition for Studies 2 and 3.* Participants in the standard condition estimated the likelihood that they would beat their competitor in each of the 14 categories (0–100%). Instructions made it clear that there would be only one winner per category, with specified tiebreakers used when needed. Participants then provided score estimates (i.e., number of correct items listed) for themselves and their competitor for each category. The order in which they made these 14 self-estimates and 14 competitor estimates was counterbalanced (as in Study 1).

*Debias condition for Studies 2 and 3.* As in Study 1, participants in the debias condition answered the score-estimate questions before providing likelihood judgments. Also, just prior to making likelihood judgments, participants read the same type of debias instructions as were used in Study 1.

## Results Not Involving the Model

Rather than reporting the findings of each study in separate sections, we report the findings from each study within the same subsections (organized by analysis or issue). This allows for useful comparisons across studies.[5] We begin by focusing on analyses that do not directly involve the EST model. Before proceeding, we should note that many analyses involve an idiographic-statistical approach in which a within-subject correlation is computed for each participant across categories, and the resulting correlations (after transformation) are then treated as data points in $t$ tests. For all such analyses reported below, Fischer $r$-to-$z$ transformations were used (e.g., see Howell, 1982).

### *Preliminary Analyses Regarding Insights About Competitive Advantage or Disadvantage*

A preliminary analysis confirmed our assumption that participants had a better sense of their comparative advantage or disadvantage in Study 2 than in Study 1, and in Study 3 than in Study 2. For each participant, we took the differences between the estimated scores for self and others (per category) and correlated those with the differences in actual self and other scores. The mean correlation from Study 1 was significantly different from zero ($r$ = .16; $p < .01$) but small in magnitude, suggesting that participants in Study 1 had only minimal insight about whether they had a competitive advantage or disadvantage for specific categories. The respective mean for Study 2 was significantly higher ($r$ = .49), $t(111) = 5.19$, $p < .001$ (for comparison of Studies 1 and 2), and it was still significantly higher for Study 3 ($r$ = .69), $t(112) = 4.14$, $p < .001$ (for comparison of Studies 2 and 3).

### *SCEs and the Influence of Debias Instructions*

To test for SCEs, we first gave each category an easiness score based on pretested participants' beliefs about how many items they

[4] In all of our experiments, participants also placed \$0.25 bets on categories of their choice (exactly half of the categories they played) after having made their set of likelihood judgments. They doubled their money if they had, in fact, won a selected category. The bets were not a main component of project; for the sake of brevity, the betting procedures and results will not be detailed in this article. However, we briefly note that the betting results generally paralleled those of likelihood judgments (although they were less sensitive to the instructional manipulation). Please contact Paul D. Windschitl for further information.

[5] Comparisons between Study 1 and either Study 2 or Study 3 must be interpreted with some caution because Study 1 was conducted separately from Studies 2 and 3 (whereas there was random assignment between Studies 2 and 3). Nevertheless, we believe that some limited across-study comparisons are useful here—given the similarities in the methodologies and participant pools that were used for the studies.

answered or listed correctly from a given category. For Study 1, these easiness scores ranged from 1.97 ($SD$ = 1.34) for baroque music to 5.53 ($SD$ = 0.89) for fast food (pretest $n$ = 64). For Studies 2 and 3, these easiness scores ranged from 0.69 ($SD$ = 0.90) for tool brands to 6.80 ($SD$ = 1.76) for planets (pretest $n$ = 31). Using an idiographic-statistical approach, we then computed—separately for each participant—the correlation between these easiness scores and his or her likelihood estimates about winning across categories. The mean of these within-subject correlations for standard and debias conditions of Studies 1–3 are listed in Table 1.

The average correlations within the standard conditions (see Table 1) were all significantly different from zero ($p$s $< .001$). These results indicate that participants in the standard condition showed the usual SCEs—expressing greater optimism about winning easy rather than hard categories. Not surprisingly, the magnitude of the SCE shrank from Study 1 (in which participants knew the least about their competitive advantage or disadvantage) to Study 3 (in which participants knew the most about their competitive advantage or disadvantage), $t(112) = 5.15$, $p < .001$.

In the debias conditions, the mean correlations were also significantly different from zero in all three studies ($p$s $< .001$). However, in Studies 1 and 2 (but not Study 3), the mean correlations were significantly lower in the debias condition than in the standard condition: Study 1, $t(54) = 2.97$, $p < .01$; Study 2, $t(58) = 2.51$, $p < .05$; Study 3, $t(56) = 0.10$, $p = .92$. In other words, in Studies 1 and 2, the SCEs were reduced but not eliminated by the debias instructions. In Study 3, the public nature of the prologue activity effectively served as a type of debiasing—making participants very aware of their competitors' strengths and weaknesses for each category. Hence, the debias instructions had no additional debiasing effect.

Figure 2A—2C provides a visual representation of the SCEs, showing how likelihood judgments shifted as a function of category easiness (from the hardest category on the left to the easiest on the right) within the standard and debias conditions. Of course, because there was exactly one winner within each pair of participants for any given category, the normative average probability of winning was 50%. Therefore, means above 50% in Figure 2A–2C reflect systematic overoptimism, and means below 50% reflect systematic overpessimism.

Table 1
*Shared-Circumstance Effect (SCE) by Study and Condition*

| | Study 1 | | Study 2 | | Study 3 | |
|---|---|---|---|---|---|---|
| Condition | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Standard | .71 | .31 | .50 | .20 | .32 | .21 |
| Debias | .42 | .40 | .33 | .28 | .32 | .24 |

*Note.* The SCE values reflect the means of the within-subject correlations between general category easiness and a person's likelihood judgments for the categories. Therefore, high means reflect that people's optimism about winning was higher for easy rather than for difficult categories. All of the means were significantly greater than zero ($ps < .001$).

## *Score Estimates and the Possible Influence of Differential Regression*

It is instructive to test whether the SCEs on the likelihood judgments are simply due to differential regression (see Chambers & Windschitl, 2004; Kruger et al., 2008; Moore, 2007; Moore & Cain, 2007; Moore & Small, 2007). According to a differential-regression account, people would predict low scores and high scores for themselves on hard categories and easy categories, respectively, but when predicting the scores of other people, all estimates would become more regressive because people have less knowledge about others. For hard categories, this would produce a negative difference between score expectations for the self and score expectations for the other (hence low optimism about winning), but for easy categories, this would produce a positive difference (hence high optimism).

To test for this possibility, we created self–other difference scores for each person by subtracting a participant's score estimates regarding the competitor from his or her score estimates for the self (for each category). Using the same idiographic-statistical approach as described above, we then computed the within-subject correlations between these self–other difference scores with the category easiness scores (see means for resulting correlations in Table 2). The mean correlations within the standard conditions were significantly different from zero in all three studies ($ps < .01$), which lends support to the idea that there was some degree of differential regression in people's score estimates. The critical comparison is between these correlations in the standard condition and the correlations from Table 1. The correlations in the two tables would be about the same if differential regression in absolute assessments accounted for the SCEs in probability judgments. However, the average correlations for the standard conditions in Table 2 were significantly and substantially smaller than those from Table 1: Study 1, $t(27) = 6.56$, $p < .001$; Study 2, $t(29) = 8.41$, $p < .001$; Study 3, $t(28) = 4.83$, $p < .001$. Hence, although differential regression might account for some portion of the SCEs involving likelihood judgments, it does not account for the entirety of those effects. In other words, it appears that some degree of differential weighting (egocentrism/focalism) must be contributing to the magnitude of the SCEs.

Figure 3A–3C provides a visual representation of these effects—showing how the self and other score estimates shifted as a function of category easiness. We collapsed across the instruction conditions for generating Figure 3A–3C, as the instruction manipulation did not significantly impact the effects.

## *The Accuracy of Likelihood Judgments and the Influence of Debias Instructions*

There are many ways of assessing the accuracy or inaccuracy of the likelihood judgments in our studies (see e.g., González-Vallejo & Bonham, 2007; Yates, 1990, 1994). We focused on three methods. First, we assessed whether participants were well calibrated—contingent on whether the categories were generally hard or generally easy. One way of indexing this type of accuracy would be to dichotomize the categories into hard and easy and then, within each of these two levels, compare the mean probability estimates with the overall percentage of wins,

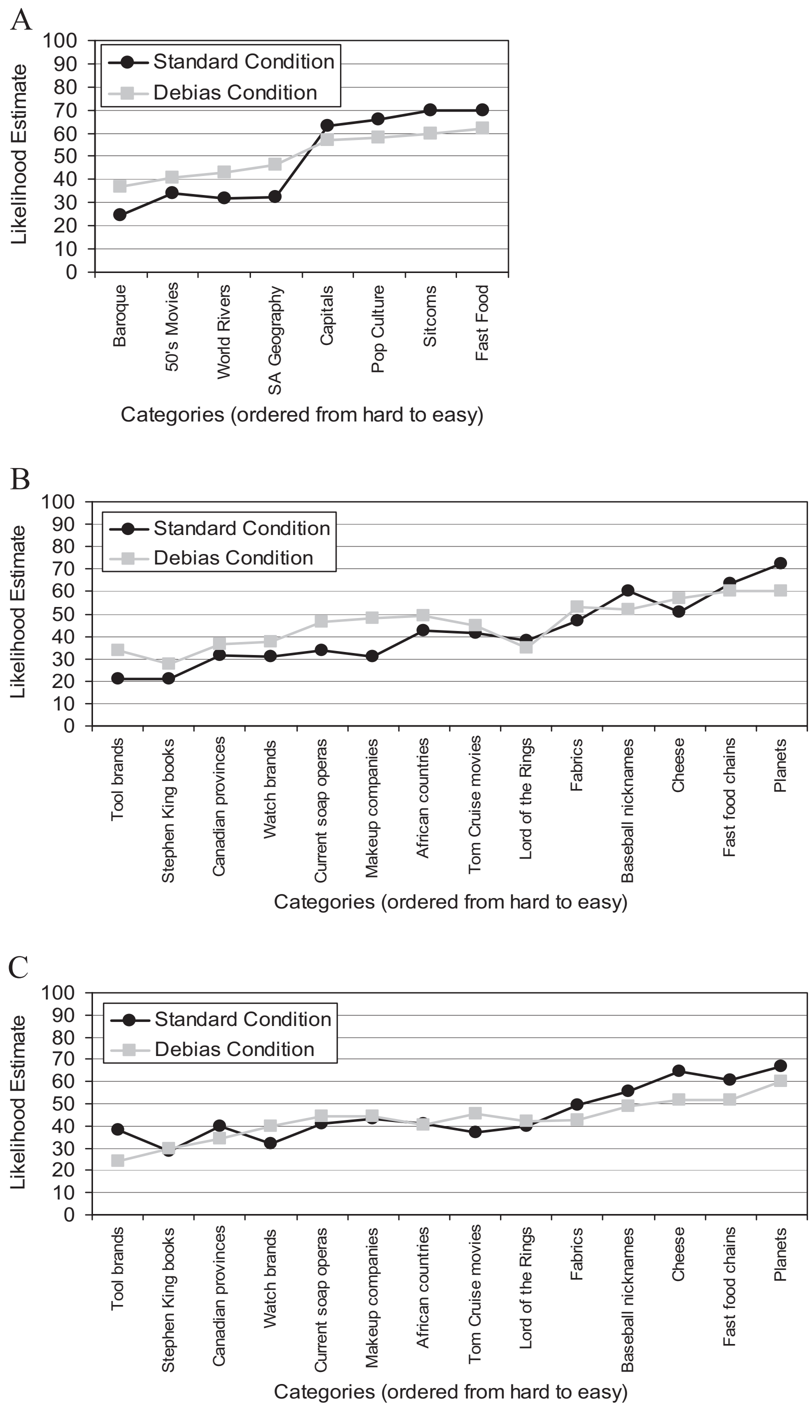


*Figure 2.* A: Mean likelihood estimates as a function of condition and category in Study 1. SA = South American. B: Mean likelihood estimates as a function of condition and category in Study 2. C: Mean likelihood estimates as a function of condition and category in Study 3.

which is necessarily 50%. However, to avoid an arbitrary dichotomization of hard and easy and to establish a single metric of this form of accuracy or inaccuracy (for group-level analyses), we can simply refer to the SCE index that was described earlier (i.e., the correlation between category easiness and likelihood judgments). To the extent that participants gave lower estimates for their probability of winning hard categories and higher estimates for their probability of winning easy catego-

Table 2
*Correlations Between Self-Other Difference Scores and Category Easiness*

| | Study 1 | | Study 2 | | Study 3 | |
|---|---|---|---|---|---|---|
| Condition | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Standard | .24 | .50 | .25 | .20 | .14 | .24 |
| Debias | .20 | .41 | .23 | .25 | .21 | .18 |

*Note.* The values reflect the means of the within-subject correlations between general category easiness and a person's self-other difference scores. Therefore, high means reflect that the self-other differences were greater for easy rather than difficult categories. All of the means were significantly greater than zero ($ps < .01$).

ries, the SCE index will have a strong positive value.[6] Hence, even though we are indexing this type of inaccuracy with a correlation, it can be considered a form of contingent, mean-level inaccuracy because it reflects systematic deviations from 50%.

Second, we assessed a form of discrimination or correlational accuracy—namely, whether people tended to give higher probability estimates to the categories they would win than to the categories they would lose. This is related to the slope component described by Yates (1990; 1994), but in our work we use a correlation as our index (i.e., the correlation between win–loss outcomes and likelihood judgments across categories).

Third, we used the mean probability score or Brier score, which is the most common global measure for gauging the accuracy of probability judgments (Brier, 1950; Yates, 1990). The Brier score includes elements of the other measures of accuracy; it is influenced by both mean-level inaccuracy and by correlational inaccuracy (i.e., poor discrimination). The Brier score for a single judgment is defined as

$$\text{Brier score} = (f - d)^2$$

where $f$ is the probability judgment and $d$ is 1 for a win and 0 for a loss. Therefore, the overall Brier score ranges from 0 to 1, with smaller numbers indicating greater accuracy.

The results for these three accuracy indexes appear in Table 3. The top two rows contain the values for the SCE index and are therefore redundant with Table 1. We have already discussed how the debias instructions improved people's judgment accuracy in the sense that they showed smaller SCEs (i.e., less of a tendency to be overpessimistic about hard categories and/or overoptimistic about easy categories). Regarding correlational or discrimination accuracy, however, Table 3 shows that this type of accuracy did not improve with the presence of debias instructions in any of the studies ($p$s = .37, .85, and .08, for Studies 1, 2, and 3, respectively). Furthermore, the Brier score also showed no statistically significant improvement (i.e., reduction) in any of the studies ($p$s > .20). A preliminary conclusion from this finding would be that the egocentrism that people use in making their likelihood judgments (which is partially removed when strong debias instructions are encountered) might not have any negative ramifications for correlational accuracy and might have few ramifications for accuracy as assessed by a standard, global accuracy measure such as the Brier score. However, we return to this issue later in the article.

## Results Involving the Model

### The Treatment of K

We now turn to analyses involving our version of the EST model, but before getting too far, we must address how we treated K within the model. Formally, K serves as a normalization constant (see Idson et al., 2001). Conceptually, K reflects a default or general expectation. As discussed earlier, norm theory might provide one perspective on what shapes K (Kahneman & Miller, 1986; see also Giladi & Klar, 2002). According to this perspective, the K for a given performance would be a norm that is partly shaped by performance representations evoked by the features of the task and partly shaped by preexisting frames of reference. This suggests that K is multiply determined. For example, when a person evaluates his or her performance on a baroque music category, the general standard (K) could be influenced by elicited representations of previous quiz performances (e.g., past performances on music quizzes). Simultaneously, K could be shaped by very recent performance experiences (the performances on the other quiz categories) and, in part, by preexisting expectations about what constitutes a general standard of good performance (e.g., a performance with 70% accuracy might be thought of as mediocre because it typically translates into a low C in college courses).

Because K is presumably shaped in so many ways, the best operationalization of K could vary somewhat across empirical contexts. We could have allowed K to vary as a free parameter—to be estimated separately for each participant in a way that maximizes model fit. Instead, however, we assumed that a participant's average self-score prediction across the categories would adequately serve as an estimate of K. That is, we assumed that self-performances on the other categories would be a highly salient context, and therefore, the self-score predictions would provide a good approximation of K. Therefore, in the modeling described below, the value of K used for a given participant was always the mean of his or her self-score predictions across all categories. We note that variations in the operationalized values of K—within plausible boundaries—would have very little impact on the modeling results that we report below because in our modeling we were solving for the values of λ that maximized correlations rather than minimized squared deviations.[7]

### Results Regarding Descriptive λ

We solved for descriptive values of λ separately for each participant. More specifically, using Equation 3 and a participant's

[6] We refer readers back to Figures 2A–2C for a depiction of how underestimations (mean responses under 50%) and overestimations (mean responses above 50%) varied across specific categories.

[7] A higher versus lower value for K would have a mean-level impact on the right side of Equation 3 (i.e., $s(A)/[s(A)+\mathrm{K}]$) and therefore on predicted probability values. However, our model maximized correlations between observed and predicted probability values, and it would therefore be mostly insensitive to such mean-level shifts. An implausibly extreme value for K (e.g., 0 or 40) would cause the right side of Equation 3 (i.e., $s(A)/[s(A)+\mathrm{K}]$) to produce similar results even when $s(A)$ varies, which would influence λ. However, as long as K is within a reasonable range (i.e., anywhere near the center of the range for $s[A]$ values), slight variations in K do not substantially impact the modeling results regarding λ.

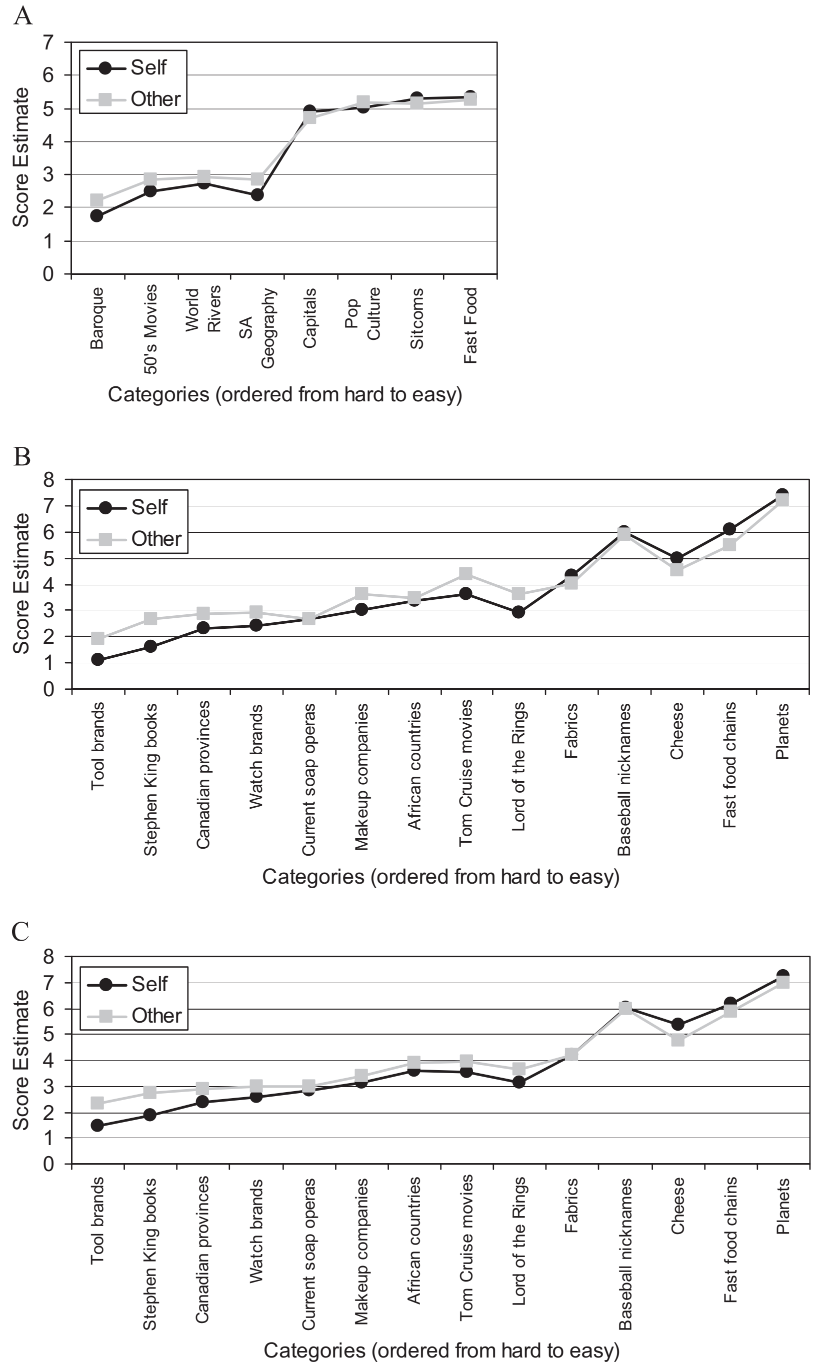


*Figure 3.* A: Mean score estimates for self and other by category in Study 1. SA = South American. B: Mean score estimates for self and other by category in Study 2. C: Mean score estimates for self and other by category in Study 3.

judgment data, we solved for the value of λ that produced model-based likelihood estimates that were maximally correlated with his or her actual likelihood judgments. The mean values for descriptive λ are shown in Table 4. Table 4 also displays the mean of the within-subject correlations between a participant's likelihood judgments and the model output (using individually estimated λs). We report the means of these correlations rather than the means of the $r^2$ values in order to preserve (when calculating a mean) the impact of instances—although rare—in which a person's judgments and model output were negatively related.

The pattern of descriptive λs in Studies 1–3 matched our expectations. The mean descriptive λ increased across Studies 1 through 3 (a one-way analysis of variance was significant), $F(2, 166) = 7.80$, $p < .001$. This is consistent with the idea that

Table 3
*Accuracy Indexes by Study and Condition*

| Index and condition | Study 1 *M* | Study 1 *SD* | Study 2 *M* | Study 2 *SD* | Study 3 *M* | Study 3 *SD* |
|---|---|---|---|---|---|---|
| Mean-level accuracy (SCE) | | | | | | |
| Standard | .71 | .31 | .50 | .20 | .32 | .21 |
| Debias | .42 | .40 | .33 | .28 | .32 | .24 |
| Correlational or discrimination accuracy | | | | | | |
| Standard | .15 | .33 | .38 | .29 | .53 | .21 |
| Debias | .06 | .34 | .35 | .27 | .43 | .23 |
| Global accuracy (Brier score) | | | | | | |
| Standard | .28 | .09 | .24 | .11 | .19 | .08 |
| Debias | .26 | .06 | .22 | .07 | .21 | .07 |

*Note.* For the SCE values, which are the same as those in Table 1, higher values reflect less accuracy. The values for correlational or discrimination accuracy reflect the means of the within-subject correlations between a person's likelihood judgments and actual win/loss outcomes across categories. Therefore, higher values reflect greater accuracy. For the Brier score, higher scores reflect less accuracy. SCE = shared circumstance effect.

egocentrism would be strongest (weakest) when participants knew the least (most) about their competitors (see e.g., Kruger et al., 2008; Windschitl et al., 2003).

As expected, the descriptive values of λ were smaller in the standard conditions than in the debias conditions for each of the three studies: Study 1, $t(53) = 3.77$, $p < .001$; Study 2, $t(55) = 5.06$, $p < .001$; and Study 3, $t(55) = 2.75$, $p < .01$. This reflects that participants who received the debias instructions became less egocentric (i.e., they tended to base their probability judgments on the comparison between the estimated scores of themselves and the estimated score of the other).

Finally, the correlations for individually fit models were generally quite high, with most means falling within the .80–.90 range (or a mean $r^2$ range of .64–.81). This indicates success in modeling people's likelihood judgments from their score estimates for themselves and their competitor.

Table 4
*Descriptive λ and Resulting Correlations for Individually Fit Models*

| Index and condition | Study 1 *M* | Study 1 *SD* | Study 2 *M* | Study 2 *SD* | Study 3 *M* | Study 3 *SD* |
|---|---|---|---|---|---|---|
| Descriptive λ | | | | | | |
| Standard | .22 | .33 | .33 | .28 | .57 | .27 |
| Debias | .59 | .39 | .70 | .27 | .76 | .25 |
| Correlations for individually fit models | | | | | | |
| Standard | .74 | .16 | .83 | .16 | .84 | .09 |
| Debias | .81 | .22 | .87 | .09 | .90 | .07 |

*Note.* Descriptive λ (ranging from 0 to 1) was estimated separately for each participant, such that it maximized the correlation between the participant's likelihood judgments and the model output. It reflects the extent to which a participant exhibited egocentrism in making likelihood judgments, with high values reflecting low egocentrism. The correlations reflect the extent to which the model output (with individually fit values of descriptive λ) correlated with participants' likelihood judgments.

### *Results Regarding Prescriptive λ*

When solving for prescriptive values of λ, we again did so separately for each participant. More specifically, using the EST equation, the score estimates for self and other, and the actual outcome values (0 = lose, 1 = win), we solved for the value of λ that produced model-based likelihood estimates that were maximally correlated with actual outcomes. The mean values for prescriptive λ are shown in Table 5. Table 5 also displays the mean of the maximized correlations (i.e., the correlations between model output and actual win–loss outcomes).

Inspection of the λs in Table 5 yields various observations. Not surprisingly, the prescriptive λs did not differ between standard and debias conditions in the three studies: Study 1, $t(53) = 1.15$, $p = .26$; Study 2, $t(54) = 0.79$, $p = .44$; and Study 3, $t(55) = 1.54$, $p = .13$. We did not expect differences in prescriptive λs because we assumed that the presence or absence of debias instructions would not impact how people should weight self- and other-assessments (just how they would weight them).

More interesting is the fact that the mean prescriptive λs fell substantially short of 1.0 within each of the three experiments (*ps* $<.01$). This result indicates that egocentric weighting—at least to some degree—would be warranted for generating optimal predictions about winning from the participants' self- and other score estimates. However, some of this prescribed egocentrism might reflect a statistical artifact. Even if the conceptually ideal value for λ was 1.0 for a given study (i.e., no egocentrism), among a sample of participants, there are bound to be instances in which idiosyncratic or chance factors within some participants' data will lead to prescriptive values of λ that are different from 1.0. Given that prescriptive λs cannot exceed 1.0, all chance-related deviations from 1.0 would fall in one direction. Therefore, the mean of the

Table 5
*Prescriptive λ and Resulting Correlations for Individually Fit Models*

| Index and Condition | Study 1 *M* | Study 1 *SD* | Study 2 *M* | Study 2 *SD* | Study 3 *M* | Study 3 *SD* |
|---|---|---|---|---|---|---|
| Prescriptive λ | | | | | | |
| Standard | .34 | .46 | .73 | .39 | .84 | .27 |
| Debias | .49 | .47 | .64 | .42 | .71 | .38 |
| Correlations (for predicting outcomes) | | | | | | |
| Standard | .29 | .32 | .45 | .23 | .58 | .16 |
| Debias | .18 | .30 | .40 | .24 | .43 | .26 |

*Note.* Prescriptive λ (ranging from 0 to 1) was estimated separately for each participant, such that it maximized the correlation between the participant's actual outcomes (0 = lose, 1 = win) and the model output. It reflects the extent to which a participant should have exhibited egocentrism in making likelihood judgments (to maximize accuracy), with high values reflecting low egocentrism. The correlations reflect the extent to which the model output (with individually fit values of prescriptive λ) correlated with a participant's actual outcomes. Conceptually, the correlations reflect how accurate a person's likelihood judgments could have been given an optimal weighting scheme (λ) and given his or her estimates of scores for the self and the competitor.

empirically calculated prescriptive λs may fall below 1.0 for statistical reasons rather than conceptually important reasons. As an illustration, even when λs are derived from actual self and other scores (predicting wins/loses), the mean of these actual score λs are below 1.0 (.79, .83, and .86 in Experiments 1, 2, and 3, respectively). These values reflect a statistical reality that some egocentric weighting of actual self and other scores would optimize the predictions of wins, but this type of egocentric weighting is not conceptually interesting and would not change the fact that nonegocentric weighting ($\lambda = 1.0$) would be the best global or a priori policy for using actual self and other scores to predict wins.

Nevertheless, there are two reasons why we can conclude that the prescriptive λs from Experiments 1–3 do indeed reflect more than a statistical artifact. First, the prescriptive λs were significantly smaller than the actual score λs from each experiment ($p <$ .001, .05, and .05 for Experiments 1, 2, and 3, respectively). A statistical-artifact account could not explain these differences. Second, mean prescriptive λ increased across Studies 1 through 3 (a one-way analysis of variance was significant), $F(2, 165) = 14.96$, $p < .001$. This pattern is consistent with our prediction that egocentrism would be less optimal as participants' knowledge about their respective competitors increased. This trend is not consistent with a statistical-artifact account. In summary, on the basis of trends in descriptive and prescriptive λs across studies, it appears that people were, and should have been, more egocentric when they knew little rather than much about their competitor.

### *Comparisons of Descriptive and Prescriptive λ*

The analyses reported in the two previous sections established that some degree of egocentrism was observed and some degree was also prescribed, but we can also examine the issue of whether the amount observed was more or less than the amount prescribed. For the standard conditions, the prescriptive λs were significantly higher than descriptive λs in Study 2, $t(29) = 5.29$, $p < .001$, and in Study 3, $t(28) = 4.86$, $p < .001$. The same trend was present but not significant for Study 1, $t(26) = 1.09$, $p = .29$. These results from Studies 2 and 3 suggest that people in the standard conditions were generally more egocentric than they should have been to maximize correlational accuracy. Within the debias conditions of all three studies, prescriptive and descriptive λs were not significantly different ($p$s $> .30$), perhaps reflecting that participants (as a group) responded well to the debias instructions.

We also compared prescriptive and descriptive λs on an individual basis to determine whether most participants would have benefited (in terms of making more accurate likelihood judgments) from exhibiting less egocentrism or more egocentrism (see Hoch, 1987, for similar analysis regarding false consensus). We classified participants into one of three groups depending on whether their descriptive λ was (a) less than their prescriptive λ—suggesting they were overly egocentric, (b) the same as their prescriptive λ—which is just right, or (c) more than their prescriptive λ—suggesting they were less egocentric than they should have been. The resulting percentages for these tallies are presented in Table 6. The results in Table 6 are consistent with those of the mean-level analyses. In the standard conditions of Studies 2 and 3, there were more participants who overdid egocentrism than underdid it: for Study 2, $\chi^2(1, N = 28) = 9.14$, $p < .01$, and for Study 3, $\chi^2(1, N = 27) = 13.37$, $p < .001$. The same trend was present but not significant for Study 1, $\chi^2(1, N = 17) = 0.53$, $p = .47$. There were no such trends in the debias conditions. In short, participants in the standard conditions of Studies 2 and 3, but not those in the debias conditions, were overly egocentric.

Table 6
*Percentages of Participants Exhibiting Too Much, Just the Right Amount of or Too Little Egocentrism*

| Condition | Study 1 | Study 2 | Study 3 |
|---|---|---|---|
| Standard | | | |
| Descript λ < Prescript. λ (too much egocentrism) | 37.0 | 73.3 | 79.3 |
| Descript. λ = Prescript. λ (just right) | 37.0 | 6.7 | 6.9 |
| Descript. λ > Prescript. λ (too little egocentrism) | 25.9 | 20.0 | 13.8 |
| Debias | | | |
| Descript. λ < Prescript. λ (too much egocentrism) | 35.7 | 40.7 | 39.3 |
| Descript. λ = Prescript. λ (just right) | 21.4 | 11.1 | 21.4 |
| Descript. λ > Prescript. λ (too little egocentrism) | 42.9 | 48.1 | 39.3 |

*Note.* The percentages were based on conditions within a given study. For example, 37.0% of participants within the standard condition of Study 1 exhibited too much egocentrism (i.e., descriptive λ that was smaller than prescriptive λ). Descript. = descriptive; Prescript = prescriptive.

### *The Consequences of Egocentrism for Accuracy*

How much did the overegocentrism exhibited by participants in the standard conditions hurt their ability to be accurate in their likelihood judgments? By using the model, we can simulate the impact that various degrees of egocentrism would or did have on accuracy. More specifically, we can compare the impact of the actually observed egocentrism with the impact of four theoretical levels of egocentrism: (a) the prescriptive level of egocentrism as established above, (b) no egocentrism, (c) moderate egocentrism, and (d) pure egocentrism. We can instantiate these theoretical conditions by setting λ equal to prescriptive λ for the first set of analyses, λ equal to 1 for the second set, λ equal to .5 for the third set, and λ equal to 0 for the fourth set. For each participant in the standard condition and for each of the simulated levels of λ, we calculated the correlation between model output and his or her actual outcomes (win or loss) across categories. Table 7 shows the means for the resulting correlations. The first row of data reflects the extent to which participants' actual probability judgments (with unaltered λs) predicted actual outcomes. The values in this row are the same as values listed for correlational or discrimination accuracy in Table 3. The next four rows reflect the results for different theoretical instantiations of λ (with the second row being identical to the correlations for the standard condition listed in Table 5). Again, the correlations in these four rows reflect the extent to which a simulated level of egocentrism yielded likelihood judgments (model output) that were accurate.

There are many possible patterns to examine from Table 7, but we guide the reader to the three we find most critical. First, the correlations tended to be slightly higher in the second row than in the first: Study 1, $t(26) = 2.31$, $p < .05$; Study 2, $t(29) = 1.83$, $p =$

Table 7
*Correlations Reflecting Actual and Possible Accuracy at Various Levels of Egocentrism in Standard Conditions of Studies 1–3*

| | Study 1 | | Study 2 | | Study 3 | |
|---|---|---|---|---|---|---|
| Accuracy | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Actual | .15 | .33 | .38 | .29 | .53 | .21 |
| When λ is set at prescribed value | .29 | .32 | .45 | .23 | .58 | .16 |
| When λ is set at 1.0 | .10 | .36 | .41 | .23 | .54 | .20 |
| When λ is set at 0.5 | .15 | .33 | .37 | .24 | .50 | .19 |
| When λ is set at 0.0 | .18 | .39 | .28 | .27 | .38 | .24 |

*Note.* The values in the first row reflect the extent to which participants' actual probability judgments correlated with actual win/loss outcomes. The values in the next four rows reflect the extent to which model output (simulating various levels of egocentrism) would correlate with actual win/loss outcomes.

.08; and Study 3, $t(28) = 1.42$, $p = .17$. This indicates that if participants had weighted their self- and other-assessments differently—using the prescriptive λ calculated by the model—they would have improved their accuracy, but only slightly. Second, the correlations in the second row were also slightly higher than those in the third row: Study 1, $t(26) = 4.12$, $p < .001$; Study 2, $t(29) = 2.57$, $p < .05$; and Study 3, $t(28) = 2.06$ $p = .05$. This indicates that the weighting specified by prescriptive λ (which involved some egocentrism) would produce slightly more accurate judgments than would a weighting scheme in which egocentrism was entirely absent. Third, except for Study 1, the correlations tended to be higher when λ was 1 than when it was 0: Study 1, $t(26) = 1.11$, $p = .28$; Study 2, $t(29) = 2.95$, $p < .01$; and Study 3, $t(28) = 3.79$, $p < .01$. This reflects that people would generally be more accurate if they exhibited no egocentrism rather than full egocentrism. The reason why Study 1 was an exception to this pattern is probably because participants knew very little about their actual comparative advantage or disadvantage. When the knowledge one holds about a competitor is not insightful about one's comparative advantage then even full use of that knowledge would not improve the accuracy of one's probability judgments.

To examine a different consequence of egocentrism, we computed the SCEs that would result when λ takes on the same theoretical values that were simulated in Table 7. Table 8 shows the magnitude of the simulated SCEs for the standard conditions. Because SCEs were indexed according to the correlations between easiness of a category and optimism about winning that category, high values reflect strong simulated SCEs.

We focus on three mains observations regarding the patterns in Table 8. First, we note that if there were absolutely no egocentrism (λ set to 1.0; see third row in Table 8), there would still be significant SCEs (Study 1, $t(26) = 3.22$, $p < .01$; Study 2, $t(29) = 8.29$, $p < .001$; Study 3, $t(28) = 3.24$ $p < .01$). These effects would be attributable to differential regression effects discussed earlier. Second, however, the SCEs would be significantly smaller than those observed from people's actual probability judgments—due to the role of the egocentrism that influenced actual probability judgments (compare first and third rows in Table 8; Study 1, $t(26) = 6.22$, $p < .001$; Study 2, $t(29) = 6.84$, $p < .001$; Study 3, $t(28) = 3.32$, $p < .01$). Third, when λs were set to their prescribed values (i.e., values optimal for achieving correlational accuracy), the simulated SCEs were large and significantly different from zero (see second row of Table 8; Study 1, $t(26) = 7.59$, $p < .001$; Study 2, $t(29) = 7.86$, $p < .001$; Study 3, $t(28) = 5.37$ $p < .001$). This is consistent with the notion that optimal weighting of information (for correlational accuracy) will yield SCEs. Thus, it appears that at least a portion of SCEs can be attributed to a sensible degree of egocentrism in the form of differential weighting (see e.g., Kruger et al., 2008). Furthermore, we note that the size of the simulated SCEs—when λ was set to the prescribed values—was greater in Study 1 than 2 ($p < .05$) and greater in Study 2 than in 3 ($p < .05$). This suggests that as people know less about their competitor, the differential weighting that optimizes their correlational accuracy will necessarily yield larger SCEs. Finally, a comparison of the last three rows of the table reveals a clear pattern in which simulated SCEs increased as the level of egocentrism increased (i.e., as λ decreased).

## Summary of the Key Findings

We collected data in three competitive environments that ranged from one in which participants had very little knowledge about the comparative strengths and weaknesses of their competitor (Study 1) to one in which participants had substantially more knowledge of their competitor (Study 3). The key findings can be summarized as follows.

1. All three studies produced robust SCEs in both the standard and debias conditions.

2. In Studies 1 and 2, the SCEs in the debias conditions were smaller than those in the standard conditions, suggesting that the debias instructions were successful in reducing people's tendency to be egocentric when gauging their optimism. Analyses of descriptive λs from our applications of the EST model confirm that participants in Studies 1 and 2 were less egocentric in the debias condition than the standard condition. For participants in Study 3, who were relatively aware of their comparative advantage or disadvantage because of the prologue activity, the debias manip-

Table 8
*Actual and Possible Shared-Circumstance Effects (SCEs) at Various Levels of Egocentrism in Standard Conditions of Studies 1–3*

| | Study 1 | | Study 2 | | Study 3 | |
|---|---|---|---|---|---|---|
| SCE | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Actual | .71 | .31 | .50 | .20 | .32 | .21 |
| When λ is set at prescribed value | .62 | .37 | .43 | .26 | .28 | .28 |
| When λ is set at 1.0 | .31 | .48 | .31 | .20 | .17 | .28 |
| When λ is set at 0.5 | .64 | .38 | .55 | .14 | .42 | .21 |
| When λ is set at 0.0 | .79 | .26 | .69 | .12 | .62 | .18 |

*Note.* The values in the first row reflect the extent to which participants' actual probability judgments correlated with general category easiness. The values in the next four rows reflect the extent to which model output (simulating various levels of egocentrism) would correlate with general category easiness.

ulation had little influence, presumably because the prologue had, de facto, already debiased all participants.

3. Based on coarse, cross-study comparisons, the magnitude of the SCEs and degree of egocentrism (highest in Study 1, lowest in Study 3) was indirectly related to the amount of knowledge participants had about their comparative advantage or disadvantage.

4. Based on prescriptive analyses—which solved for the level of $\lambda$ that would maximize the correlational accuracy—the highest levels of egocentrism were prescribed in Study 1 (in which knowledge about one's competitor was low) and the lowest in Study 3 (in which knowledge about one's competitor was higher).

5. A comparison of descriptive and prescriptive $\lambda$s revealed that most participants in the standard conditions of Studies 2 and 3 would have been slightly more accurate in their likelihood judgments (in terms of giving likelihood judgments that correlated with actual wins or losses) if they had been less egocentric than they were. This trend was not significant for Study 1 because participants had so little information about their competitor that being nonegocentric did not necessarily result in using more reliable information.

6. By using the EST model and plugging in theoretical values for $\lambda$, we learned more about the consequences of egocentrism. Relative to when we simulated the complete absence of egocentrism ($\lambda = 1$), the simulation of full egocentrism in standard conditions ($\lambda = 0$) led to slightly poorer correlational accuracy in likelihood judgments (Studies 2 and 3) and substantially poorer mean-level accuracy in the form of larger SCEs (Studies 1, 2, and 3).

7. Finally, although the debias instructions generally reduced SCEs (as mentioned in Point 2 above), the debias instructions did not prompt greater accuracy in terms of correlations or on the Brier score.

## General Discussion

Our work illustrates the benefits of taking an EST approach to understanding people's optimism in competitions. The work also reveals a host of information about whether egocentrism helped or hurt accuracy and how people respond to debiasing instructions regarding egocentrism. We start our General Discussion with an examination of these two topics. Then we discuss how EST could be fruitfully applied to other judgments, such as comparative ability or trait judgments. We also discuss the relationship between EST and other accounts of comparative bias, and before concluding, we explain the benefits of using EST over conventional methods of assessing egocentrism in comparative judgments.

### *Is Egocentrism a Good Thing? Are People Excessive or Judicious in Their Egocentrism?*

The verdict on whether egocentrism was a good thing in these studies is complex and depends on what type of accuracy serves as the gold standard. Regarding mean-level accuracy as indexed by the SCE, egocentrism was always bad. The SCEs simulated by the model were smallest when egocentrism was completely absent rather than partially or fully present (see bottom three rows of Table 8). Also, the SCEs that were actually observed would have been smaller if no egocentrism were present (compare Rows 1 and 3 of Table 8). Regarding correlational accuracy, the short answer as to whether egocentrism is a good thing is as follows: small levels of egocentrism can be good, but people tend to overdo egocentrism (in the standard conditions). The prescriptive $\lambda$s tended to prescribe some degree of egocentrism, and models involving prescriptive $\lambda$s produced slightly more accurate judgments than did models simulating no egocentrism (compare Rows 2 and 3 of Table 7). Yet, for most participants in the standard conditions of Studies 2 and 3, the accuracy of their likelihood judgments would have been better if they were less rather than more egocentric. The extent to which they overdid egocentrism, however, caused only mild damage to the correlation accuracy. To summarize, egocentrism at any degree is bad for mean-level accuracy, and although egocentrism has potential to help correlational accuracy, the degree of egocentrism that people tend to show outpaces the level that would be optimal for correlation accuracy. Overall—considering both mean-level and correlational accuracy—it appears that people in competitions like these would be wise to be less egocentric than they are normally inclined to be.

Although we have just noted that our participants were generally excessive in their egocentrism, we also note that there were signs of judiciousness. Most notably, our prescriptive analyses suggested that participants in Study 3 should be the least egocentric, whereas those in Study 1 should be the most egocentric. Indeed, the analyses of descriptive $\lambda$s revealed that this was the observed pattern across Studies 1–3.

Our findings regarding egocentrism constitute an important extension beyond recent articles that have noted the possibility of a rational grounding for some egocentrism or differential weighting (see Burson & Klayman, 2006; Chambers & Windschitl, 2004; Kruger et al., 2008; see also Moore & Small, 2007). As discussed earlier, these articles suggest that because people tend to have more knowledge about themselves than about others, any assessments they make of themselves would tend to be more valid than assessments they make of others; therefore, people might have good reason for giving more weight to assessments about the self than to assessments about another. Of these articles, only one study from Burson and Klayman (2006), which is discussed below, empirically assessed what the optimal weighting of self and other-assessments would be. Therefore, the current article is notable in being one of the first to test whether egocentrism is, in fact, a good thing. Whereas previous articles speculated that egocentrism might be beneficial for correlational accuracy (and perhaps a composite form of overall accuracy), our results provide more detailed and empirically backed conclusions. Namely, egocentrism is indeed potentially useful for maximizing correlational accuracy, but the level of egocentrism that people tend to exhibit is too extreme.

This conclusion can be compared with the conclusion that Burson and Klayman (2006) drew from their experiment—an experiment that was quite different from our own. Participants engaged in word prospector tasks, and the main dependent variable was a percentile judgment of how one's performance compared with the performance of all other students rather than a likelihood judgment about beating a single competitor. Burson and Klayman manipulated whether people received direct feedback about their own performance and the median performance of others. On the basis of regression analyses, which treated actual percentile as the criterion variable, they concluded that their participants' relative weighting of self-assessments and other-assessments should have been sensitive—but were not sensitive—to the feedback manipulations, which suggests that people are insensitive to the diagnosticity of self- and other-assessments. This conclusion is at odds with our finding that as people knew more about their competitor (more in Experiment 3 than in 2 than in 1), they were less egocentric.[8] However, because there are so many distinctions between the methodologies of our studies and that of the study by Burson and Klayman, we do not attempt to discuss all the possible reasons for the different conclusions. Clearly, this would be a good area for further research.

Our conclusions can also be compared with the conclusions drawn by researchers looking at a somewhat parallel issue from the false consensus literature (Davis et al., 1986; Dawes & Mulford, 1996; Hoch, 1987, 1988; Krueger & Clement, 1994; Krueger & Zeiger, 1993). Generally, a false consensus effect is said to occur when people overestimate the extent to which their own characteristics, attitudes, or behaviors are shared by others (Marks & Miller, 1987; Mullen et al., 1985). A particular false consensus effect might be manifested as follows: Relative to participants who do not endorse a particular attitude, those who do endorse the attitude give higher estimates of the general prevalence of that attitude (e.g., Ross et al., 1977). Davis, Hoch, Dawes, and their respective coauthors (Davis et al., 1986; Dawes & Mulford, 1996; Hoch, 1987, 1988) argued that although the false consensus effect might appear to suggest faulty reasoning or beliefs on the part of respondents, the effect is actually the result of a sensible projection. From a Bayesian perspective, the self provides a useful data point for estimating a population statistic (see Dawes & Mulford, 1996). A related point is that when people have good information about the self and only sketchy or unreliable information about others, using the self as an anchor for projecting about others might be a sensible strategy if there is some actual similarity between the self and the target being estimated (Hoch, 1987). Hoch (1987) used a modeling and data analysis strategy that has parallels with ours, such as the computation of a variable reflecting the weight that self-characteristics should receive relative to other information. Using this approach, he demonstrated that projection could be useful and was generally not overused for optimizing accuracy in people's estimates about others (but for an alternative perspective, see Krueger & Clement, 1994; Krueger & Zeiger, 1993). Although Hoch's (1987) conclusions might seem to be at odds with the conclusions we generated from our experiments, there is no real conflict. The false consensus research of Hoch (1987) and others focused on how projection influences the accuracy of people's estimates about others, whereas our work focuses on how egocentrism influences the accuracy of likelihood judgments about competition outcomes.

### *The Influence of Debias Instructions*

As already reported, the debias instructions had mixed influences on accuracy. Generally speaking, they led to greater mean-level accuracy (i.e., smaller SCEs in Studies 1 and 2), even though they had little influence on correlational accuracy. What accounts for this apparent inconsistency? We suggest that when people are strenuously urged to avoid egocentrism (as we did in our debias conditions), people are quite capable of shifting additional attention and weight to their competitor's strengths and weaknesses. Indeed, the descriptive $\lambda$ values from Studies 1 and 2 reflect less egocentrism in the debias conditions. This reduction in egocentrism—as any reduction in egocentrism would be—was successful in reducing the magnitude of SCEs. However, this attentional shift, we suspect, was rather crude. Akin to throwing the baby out with the bath water, participants' shifts of attention may have offset egocentrism that was not rationally grounded (e.g., a chronic attentional bias) as well as egocentrism that was rationally grounded (i.e., a tendency to weight reliable assessments more heavily than less reliable assessments). The result was that people were no more accurate, in a correlational sense, with debiasing than without. Being more accurate in a correlational sense would require that people not only react to the debias instruction but also react in a way that preserved any helpful bias yet removed any unhelpful bias. This is generally a large hurdle for debiasing attempts (see e.g., Larrick, 2004; Wegener & Petty, 1997; see also Nisbett & Wilson, 1977).

### *EST as an Integrative Framework*

Thus far, we have discussed and shown how our application of EST can be quite useful for conceptualizing, measuring, and simulating the effects of egocentrism on optimism in competitions. However, EST has broad potential and can be applied to understanding bias in many types of referent-dependent judgments (i.e., judgments that require evidence for a target to be compared with evidence for a specific referent or set of referents; see Windschitl et al., 2008). These referent-dependent judgments include comparative ability and trait judgments (e.g., Alicke & Govorun, 2005; Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Dunning, Meyerowitz, & Holzberg, 1989; Goethals, Messick, & Allison, 1991; Hoorens, 1995; Kruger, 1999; Pahl & Eiser, 2005), comparative optimism judgments (e.g., Blanton et al., 2001; Burger & Burns, 1988; Chambers et al., 2003; Eiser et al., 2001; Heine & Lehman, 1995; Helweg-Larsen & Shepperd, 2001; Klar, Medding, & Sarel, 1996; Klein & Weinstein, 1997; Price, 2001; Price, Smith, & Lench, 2006; Rose, Endo, Windschitl, & Suls, in press; Weinstein, 1980, 1984), and generic direct-comparison judgments (e.g., Giladi & Klar, 2002; Klar, 2002; Posavac, Brakus, Jain, & Cronley, 2006; Suls, Krizan, Chambers, & Mortensen, 2007; Windschitl et al., 2008).

The application of EST to comparative ability judgments (e.g., "Relative to the average person, how good am I at dancing?") is straightforward and differs little from how we applied EST to likelihood judgments about winning. The output of the model would merely need an additional transformation, if desired, so that answers fall on a smaller range (e.g., $-3$ to $+3$) rather than 0–100%.

In applying EST to comparative optimism judgments (e.g., "Relative to other people, how likely are you to acquire skin cancer?"), $s(A)$ and $s(B)$ would reflect the subjective absolute likelihoods of the self and others experiencing the specified event. K would be the default subjective likelihood for events in general. If the event in question is one that is generally frequent or likely,

[8] The findings from Burson and Klayman (2006) that are most relevant to our work involve the cell in their design in which participants received no direct feedback. However, the results from this cell are difficult to interpret, perhaps in part because participants in that cell did not have a very good sense of their relative standing (much like our Experiment 1). This lack of knowledge caused the optimal weights for the predictors in Burson and Klayman's regression analysis to be close to zero.

then $s(A)$ would tend to be greater than K, and egocentric respondents would tend to indicate that they are more likely than others to experience the event (which would represent comparative optimism or pessimism depending on whether the event is positive or negative in valance; see Chambers et al., 2003; Kruger & Burrus, 2004).

In applying EST to generic direct-comparison judgments ("Relative to the other hotels in the set, how desirable is this hotel?"), $s(A)$ and $s(B)$ would reflect the absolute assessments regarding the focal and referent items. K would be a default or general expectancy for the types of items being considered. EST would predict that individual focal items drawn from an attractive or otherwise superior set of items would tend to be rated as comparatively better than the other items in the set because $s(A)$ would be greater than K. Focal items from an inferior set would be rated as comparatively worse because $s(A) < K$ (see e.g., Giladi & Klar, 2002; Suls et al., 2007; Windschitl et al., 2008).

The fact that EST can be applied to a wide range of referent-dependent judgments is important. Until very recently, there has been relatively little contact between the research literature concerning basic probability judgments and the research literature concerning comparative-ability, comparative-optimism, and generic-comparative judgments. This is unfortunate given that all of these types of judgments are referent-dependent and therefore appear to have some structural similarities (yet also differences).

We make no claim that EST is the only way to model the structural similarities shared by referent-dependent judgments. However, we do believe that EST is an excellent place to start. As the name reflects, EST is an extension of support theory (Tversky & Koehler, 1994), whose notion of subadditivity is useful for explaining some very basic and robust characteristics of probability judgments. We did not delve into issues regarding subadditivity here; subadditivity becomes relevant when there is more than one alternative to a focal hypothesis or event. However, support theory—and therefore EST—provides a formalized and already tested foundation for future research that might involve explicitly investigating situations relevant to subadditivity (i.e., assessing evidence for multiple rather than single alternatives to a focal hypothesis).

Another related conceptualization that could be used as a global framework for referent-dependent judgments is the local standard and general standard (LOGE; Giladi & Klar, 2002) concept. In their LOGE conceptualization, Giladi and Klar (2002) hypothesized that when people are asked to make direct comparison judgments, they should compare the focal item (or person) exclusively with a local standard—namely the referent items specified by the question. However, in part, people compare the focal item with a general standard, such as one based on all items of its type in memory. Thus, LOGE and EST are broadly similar, with the notion of a general standard in LOGE being almost the same as K in EST. However, LOGE was not articulated as a formalized model. As such, it is not readily equipped to handle probability judgments and subadditivity, nor does it have a convenient way of quantifying the specific weight given to local and general standards. EST is articulated as a specific mathematical equation and contains the $\lambda$ weighting factor. Therefore, between EST and LOGE, which again share core features, we find EST to be more useful for examining specific results from referent-dependent judgments and for conceptually integrating various forms of referent-dependent judgments.

### *Does EST Replace Egocentrism and Other Differential Weighting Accounts?*

Our application of EST constitutes an improvement for researchers' conceptualization of how biases such as egocentrism can result in SCEs, above- and below-average effects, and other effects. However, EST in no way replaces or supersedes those accounts. This is true for two related reasons. First, EST is not intended as a literal process account. When analyses on a study's dataset suggest that $\lambda$ was less than 1.0, there are several cognitive process explanations for this result. The explanation that seems most tied to the EST equation is that $\lambda$ is less than 1.0 because respondents first made separate absolute judgments about the self and their competitor, but when generating their likelihood judgment they partially compared their absolute judgment of the self with a general standard. A related but different explanation would suggest that people first thought about whether the task was something that they themselves were good at, developing an initial sense of optimism or pessimism, then adjusted this optimism—but insufficiently—based on thoughts about whether the task was something that the competitor would be good at (see the anchoring explanation by Kruger, 1999). Yet another possibility is that people strategically gave more weight to evaluations in which they had more confidence rather than less confidence. These accounts (and others; see Chambers & Windschitl, 2004) involve different processes, but all can be represented by the model because the key factor in each is the extent to which people's likelihood judgments are influenced by assessments regarding their competitor.

A second and related reason for why EST does not replace specific egocentrism and related accounts is that the EST model is agnostic about the elements of the judgment context that are the preconditions for differential weighting (see Windschitl et al., 2008). That is, the structure of the model would be identical for representing the differential weighting that is due to generic focalism (i.e., the fact that one entity was denoted as focal in the question that solicited a response), some form of egocentrism (i.e., the fact that the question asked about the self), or some difficulty in assessing support for a group of competitors (i.e., the fact that respondent faced more than one competitor).

Consequently, instead of being viewed as a replacement for any account of differential weighting, it should be viewed as a useful formalization of the impact that knowledge about a focal entity and a referent has on a referent-dependent judgment. This formalization, in contrast to verbal accounts, allows for a quantification of differential weighting that is relatively interpretable and precise. It also is amenable to both descriptive and prescriptive analyses.

### *EST and the Regression (Path Analysis) Approach to Assessing Differential Weighting*

Another advantage of the EST approach is that it avoids problems that often plague the conventional path analysis approach, which has been reported in numerous recent articles (e.g., Chambers & Suls, 2007; Chambers et al., 2003; Eiser et al., 2001; Giladi & Klar, 2002; Klar, 2002; Klar & Giladi, 1997, 1999; Kruger, 1999; Kruger & Burrus, 2004; Kruger et al., 2008; Moore & Kim, 2003; Windschitl et al., 2003). In a typical application of the path analysis, researchers conduct regression analyses in which absolute judgments of the self and the other (e.g., the average peer, a friend, a competitor) are entered as predictor variables, whereas comparative or likelihood judgments serve as the criterion variable. A common finding that is often interpreted as support for differential weighting is that the beta weight for self is positive and strong, whereas the beta weight for other is near zero.[9]

One problem that sometimes plagues the interpretation of these path analyses is relevant to cases in which the variability across evaluations of the other is smaller than the variability across evaluations of the self. In such cases, the beta for the self can be larger than the beta for the other, even when both self and other are equally considered when people formulate a comparative or likelihood judgment. The problem is not that path or regression analyses tend to reward higher betas to predictors with more variability. In fact, regressions have the useful property of controlling for the differential variability. Instead, the problem is related to what happens when people formulate their comparative judgment. If they make a comparison between the self and the other—whether it be a difference comparison (self – other), a ratio comparison (self / other), or a proportion comparison [self / (self + other)]—they are fully considering both the self and the other. There is no differential weighting. However, across participants or across items, the variability in the numeric results of these comparisons will necessarily be driven more by the variable with high variability (self) than the one with low variability (other). Hence, even if we ran a regression with the numeric results of these comparisons as the criterion (say self – other), the beta for the self would be larger than the beta for the other. Therefore, in short, in cases in which the beta for the self is greater than the beta for the other, but the variability in the self is also greater than the variability in the other, it is then premature to conclude that there was differential weighting.

Our use of EST does not suffer from this problem because, in using the EST model, we are testing the extent to which a given judgment model is a good one for representing the comparative judgment. If a comparison between the self and the other was critical in shaping people's comparative judgments, this would be reflected in a strong weight (high $\lambda$) for the respective component of the model [self / (self + other)], regardless of whether the self contained more variability than the other.

A related point is that on a practical level, the value of $\lambda$ provides a clearer and more useful way of measuring differential weighting than does a comparison of the beta weights from a regression involving the self and the other. The value of $\lambda$ ranges from 0 to 1.0 and can serve as a unitary index of differential weighting. Alternatively, in the conventional path analysis approach, betas come in pairs and do not have known ranges. The researcher must compare the $\beta_{SELF}$ with the $\beta_{OTHER}$ to infer something about the relative contribution of the self and the other. But this comparison is not always easily interpreted, such as when a $\beta_{OTHER}$ is on the positive side of zero, even though it presumably should be on the negative side. Complicating matters further is the fact that $\beta_{SELF}$ and $\beta_{OTHER}$ can be systematically influenced by the predictive validity of the overall model (e.g., by the level of noise in comparative judgments). Alternatively, $\lambda$ is isolated as a weighting parameter whose value does not rise and fall as a function of the predictive validity of the overall model (although it will become less stable as noise in comparative judgments increases).

We wish to emphasize that although there are problems that can plague the conventional path analysis approach (as described here and elsewhere; see Moore, 2007), these problems do not affect all such analyses (see Rose & Windschitl, 2008, for discussion). For example, when the referent is one person rather than a group, the variability in the self is not always greater than in the other (e.g., Windschitl et al., 2003). Also, although we have suggested that interpretations of differential weighting based on $\beta_{SELF}$ and $\beta_{OTHER}$ comparisons can be rather challenging, they have nevertheless been quite useful in detecting coarse differences or changes in differential weighting across studies or conditions (e.g., see Eiser et al., 2001; Giladi & Klar, 2002; Krizan & Windschitl, 2007; Kruger et al., 2008). Finally, it is notable that in large measure, a conventional, path-analytic approach to our data, which we do not report here, suggested similar conclusions about differential weighting and its changes across experiments, as did our EST approach. In short, conventional path analyses can be useful under the right conditions, but the EST approach offers improvements in terms of interpretability and precision.

## Conclusion

In conclusion, EST can serve as a useful tool for understanding a variety of types of referent-dependent judgments. In the present work, EST was specifically used to explore the impact of egocentrism on the accuracy of people's optimism. We found that although egocentric weighting of information does have some conceivable benefits (for correlational accuracy), people tended to overdo egocentric weighting. When urged to avoid it, they reduced the egocentrism that was implicit in their reported optimism. This improved their mean-level accuracy (reduced SCEs) but had no net impact on their correlational accuracy. At a practical level, then, there appears to be no downside to urging people to avoid being egocentric. By extension, these results suggest that the high levels of egocentrism that people exhibit in social-comparative ability judgments (e.g., Kruger, 1999) and comparative optimism judgments (Chambers et al., 2003; Kruger & Burrus, 2004; Weinstein & Lachendro, 1982) hurt rather than help the overall accuracy of those judgments. With that said, however, we recognize that we cannot rule out the possibility that egocentrism's benefits outweigh its liabilities within some—as yet unidentified—contexts. For future work exploring such contexts, EST and the analytic strategies described here offer a useful framework.

[9] Other researchers have recently noted caveats that could apply to some studies using the path analysis approach. For example, if there is any form of conflation of the absolute self-judgments and comparative judgments, or if self-judgments can be based on comparative evaluations, then the beta for self will necessarily be high (see Burson & Klayman, 2006; Moore & Cain, 2007; Moore & Small, 2007). We note that neither of these issues applies to our study because the absolute questions were quite concrete and were answered on numeric, common-rule scales.

## References

Alicke, M. D., & Govorun, O. (2005). The better-than-average-effect. In M. D. Alicke, D. A. Dunning, & J. I. Krueger (Eds.), *The self in social judgment* (pp. 85–106). Philadelphia: Psychology Press.

Alicke, M. D., Klotz, M. L., Breitenbrecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the above-average effect. *Journal of Personality and Social Psychology, 68,* 804–825.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39,* 1037–1049.

Blanton, H., Axsom, D., McClive, K., & 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,* 1627–1636.

Brenner, L. A., Koehler, D. J., & Rottenstreich, Y. (2002). Remarks on support theory: Recent advances and future directions. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 489–509). New York: Cambridge University Press.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78,* 1–3.

Burger, J. M., & Burns, L. (1988). The illusion of unique invulnerability and the use of effective contraception. *Personality and Social Psychology Bulletin, 14,* 264–270.

Burson, K. A., & Klayman, J. (2006). *Judgments of performance: The relative, the absolute, and the in between.* Unpublished manuscript.

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

Chambers, J. R., & Suls, J. (2007). The role of egocentrism and focalism in the emotion intensity bias. *Journal of Experimental Social Psychology, 43,* 618–625.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130,* 813–838.

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,* 1343–1356.

Davis, H. L., Hoch, S. J., & Ragsdale, E. K. E. (1986). An anchoring and adjustment model of spousal prediction. *Journal of Consumer Research, 13,* 25–37.

Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment of flaws in how we study judgment. *Organizational Behavior and Human Decision Processes, 65,* 201–211.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082–1090.

Eiser, J. R., Pahl, S., & Prins, Y. R. A. (2001). Optimism, pessimism, and the direction of self-other comparisons. *Journal of Experimental Social Psychology, 37,* 77–84.

Endo, Y. (2007). Optimistic and pessimistic biases and comparative judgmental processes in Japan: Do people really compare themselves to their peers? *Asian Journal of Social Psychology, 10,* 224–232.

Endo, Y. (2008). The role of egocentrism in judgmental biases in social comparison under shared circumstances. *The Japanese Journal of Experimental Social Psychology, 47,* 134–145.

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

Fox, C. R., & Levav, J. (2000). Familiarity bias and belief reversal in relative likelihood judgment. *Organizational Behavior and Human Decision Processes, 82,* 268–292.

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,* 538–551.

Goethals, G. R., Messick, D. M., & Allison, S. T. (1991). The uniqueness bias: Studies of constructive social comparison. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research.* (pp. 149–176). Hillsdale, NJ: Erlbaum.

González-Vallejo, C., & Bonham, A. (2007). Aligning confidence with accuracy: Revisiting the role of feedback. *Acta Psychologica, 125,* 221–239.

Heine, S. J., & Lehman, D. R. (1995). Cultural variation in unrealistic optimism: Does the West feel more invulnerable than the East? *Journal of Personality and Social Psychology, 4,* 595–607.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review, 51,* 74–95.

Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology, 53,* 221–234.

Hoch, S. J. (1988). Who do we know? Predicting the interests and opinions of the American consumer. *Journal of Consumer Research, 15,* 315–324.

Hoorens, V. (1995). Self-favoring biases, self-presentation, and the self-other asymmetry in social comparison. *Journal of Personality, 63,* 793–817.

Howell, D. C. (1982). *Statistical methods for psychology (2nd ed.).* Boston: Duxbury Press.

Idson, L. C., Krantz, D. H., Osherson, D., & Bonini, N. (2001). The relation between probability and evidence judgment: An extension of support theory. *Journal of Risk and Uncertainty, 22,* 227–249.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

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

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,* 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,* 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, 5,* 585–594.

Klar, Y., Medding, A., & Sarel, D. (1996). Nonunique invulnerability: Singular versus distributional probabilities and unrealistic optimism in comparative risk judgments. *Organizational Behavior and Human Decision Processes, 67,* 229–245.

Klein, W. M., & Weinstein, N. D. (1997). Social comparison and unrealistic optimism about personal risk. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well–being: Perspectives from social comparison theory* (pp. 25–61). Hillsdale, NJ: Erlbaum.

Koehler, D., Brenner, L., & Tversky, A. (1997). The enhancement effect in probability judgment. *Journal of Behavioral Decision Making, 10,* 293–313.

Krizan, Z., & Windschitl, P. D. (2007). Team allegiance can lead to both optimistic and pessimistic predictions. *Journal of Experimental Social Psychology, 43,* 327–333.

Krueger, J. (2003). Return of the ego—Self-referent information as a filter for social prediction: Comment on Karniol (2003). *Psychological Review, 110,* 585–590.

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

Krueger, J., & Zeiger J. S. (1993). Social categorization and the truly false consensus effect. *Journal of Personality and Social Psychology, 65,* 670–680.

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

Kruger, J., Windschitl, P. D., Burrus, J., Fessel, F., & Chambers, J. R. (2008). On the rational side of egocentrism in social comparisons. *Journal of Experimental Social Psychology, 44,* 220–232.

Larrick, R. (2004). Debiasing. In Koehler, D. J. & Harvey, N. (Eds.), *Blackwell handbook of judgment and decision making* (pp. 317–337). Malden, MA: Blackwell Publishers.

Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). *Organizational Behavior and Human Decision Processes, 102,* 76–94.

Lehman, D. R., Krosnick, J. A., West, R. L., & Li, F. (1992). The focus of judgment effect: A question wording effect due to hypothesis confirmation bias. *Personality and Social Psychology Bulletin, 18,* 690–699.

Macchi, L., Osherson, D., Krantz, D. H. (1999). A note on superadditive probability judgment. *Psychological Review, 106,* 210–214.

Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin, 102,* 72–90.

McKenzie, C. R. M. (1998). Taking into account the strength of an alternative hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 771–792.

Moore, D. A. (2007). 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, 102,* 42–58.

Moore, D. A., & Cain, D. M. (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes, 103,* 197–213.

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

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,* 972–989.

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderkiok, M. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology, 21,* 262–283.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Pahl, S., & Eiser, J. R. (2005). Valence, comparison focus, and self-positivity biases: Does it matter whether people judge positive or negative traits? *Experimental Psychology, 52,* 303–310.

Pahl, S., & Eiser, J. R. (2007). How malleable is comparative self-positivity? The effects of manipulating judgmental focus and accessibility. *European Journal of Social Psychology, 37,* 617–627.

Posavac, S. S., Brakus, J. J., Jain, S. P., & Cronley, M. L. (2006). Selective assessment and positivity bias in environmental valuation. *Journal of Experimental Psychology: Applied, 12,* 43–49.

Price, P. C. (2001). A group size effect on personal risk judgments: Implications for unrealistic optimism. *Memory & Cognition, 29,* 578–586.

Price, P. C., Smith, A. R., & Lench, H. C. (2006). The effect of target group size on risk judgments and comparative optimism: The more, the riskier. *Journal of Personality and Social Psychology, 90,* 382–398.

Rose, J. P., Endo, Y., Windschitl, P. D., & Suls, J. (in press). Cultural differences in unrealistic optimism and pessimism: The role of egocentrism and direct versus indirect comparison measures. *Personality and Social Psychology Bulletin.*

Rose, J. P. & Windschitl, P. D. (2008). How egocentrism and optimism change in response to feedback in repeated competitions. *Organizational Behavior and Human Decision Processes, 105,* 201–220.

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,* 279–301.

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

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review, 104,* 406–415.

Sherman, S. J., Presson, C. C., & Chassin, L. (1984). Mechanisms underlying the false consensus effect: The special role of threats to the self. *Personality and Social Psychology Bulletin, 10,* 127–138.

Suls, J., Krizan, Z., Chambers, J. R., & Mortensen, C. (2007). The role of focalism and group diffuseness in comparative biases. Manuscript submitted for publication.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101,* 547–567.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naïve theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 141–208). New York: Academic Press.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39,* 806–820.

Weinstein, N. D. (1984). Why it won't happen to me: Perceptions of risk factors and susceptibility. *Health Psychology, 3,* 431–457.

Weinstein, N. D., & Klein, W. M. (1995). Resistance of personal risk perceptions to debiasing interventions. *Health Psychology, 14,* 132–140.

Weinstein, N. D., & Lachendro, E. (1982). Egocentrism as a source of unrealistic optimism. *Personality and Social Psychology Bulletin, 8,* 195–200.

Wilson, T. D., & Brekke, N. C. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.

Windschitl, P. D. (2000). The binary additivity of subjective probability does not indicate the binary complementarity of perceived certainty. *Organizational Behavior and Human Decision Processes, 81,* 195–225.

Windschitl, P. D., Conybeare, D., & Krizan, Z. (2008). Direct-comparison judgments: When and why above- and below-average effects reverse. *Journal of Experimental Psychology: General, 137,* 182–200.

Windschitl, P. D., Kruger, J., & Simms, E. N. (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,* 389–408.

Yamagishi, K. (2002). Proximity, compatibility, and noncomplementarity in subjective probability. *Organizational Behavior and Human Decision Processes, 87,* 136–155.

Yates, J. F. (1990). *Judgment and decision making.* Englewood Cliffs, NJ: Prentice Hall.

Yates, J. F. (1994). Subjective probability accuracy analysis. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 381–410). Chichester, England: Wiley.

# Appendix

## Debias Instructions

Below are the debias instructions that were used in Study 1. The debias instructions for Studies 2 and 3 were nearly identical to those for Study 1, except that the instructions for Studies 2 and 3 referred to items that were listed rather than questions that were answered correctly.

Before you begin making likelihood judgments, we need to inform you about a bias that affects how people think about their likelihood of winning a competition. Winning a particular category depends on two things:

1. The number of questions you answered correctly.

2. The number of questions your competitor answered correctly.

Both 1 and 2 are equally important for determining the winner.

You have just indicated your projections for both of these values, so keep these in mind when making likelihood judgments. Previous research has shown that when people are thinking about their likelihood of winning a category, they mistakenly tend to think only about how many questions they've answered correctly and not how many questions the other person probably answered correctly. This is a biased form of thinking that we hope you will avoid when judging your likelihood of winning categories.

Therefore, before giving each likelihood estimate, first consider both your performance and your competitor's performance. Only then should you respond.

Again, don't forget to consider how well or poorly your competitor probably did on a category before you give your likelihood estimate.

Received September 10, 2007
Revision received March 7, 2008
Accepted March 16, 2008 ■