Available online at www.sciencedirect.com
ScienceDirect
Organizational Behavior and Human Decision Processes 105 (2008) 201–220
ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
www.elsevier.com/locate/obhdp

How egocentrism and optimism change in response to feedback in repeated competitions

Jason P. Rose *, Paul D. Windschitl

Department of Psychology, University of Iowa, Iowa City, IA 52242, USA

Received 4 August 2006
Available online 30 October 2007
Accepted by Robyn Dawes

Abstract

People tend to egocentrically focus on how adverse or beneficial conditions in competitions affect the self, while inadequately considering the comparable impact on opponents. This leads to overoptimism for a victory in easy tasks and underoptimism in hard tasks. Four experiments investigated whether experience and performance feedback in a multi-round competition would influence egocentric weighting and optimism biases across rounds. The results indicated that egocentric weighting and optimism biases decreased across rounds. However, this apparent debiasing occurred under restrictive conditions, and participants did not generalize their learned, non-egocentric tendencies to novel contexts. The roles of differential confidence and surface/structural similarity are discussed as reasons why optimism biases were generally pervasive.

Keywords: Egocentrism; Shared-circumstance effect; Comparative judgment; Subjective probability; Overoptimism; Underoptimism; Learning; Feedback; Experience

Introduction

People appear to be generally egocentric when estimating their likelihood of winning a competition (Moore & Kim, 2003; Windschitl, Kruger, & Simms, 2003). They tend to think about whether the circumstances of the competition are favorable or unfavorable for their own performance, while giving less consideration to whether the circumstances are favorable or unfavorable for the performance(s) of their competitor(s). This egocentrism can result in overoptimism when the task in the competition is generally easy or the circumstances of the competition are generally favorable for strong performances from all competitors. However, egocentrism can result in underoptimism when the task is generally difficult or the circumstances are unfavorable. For example, when college student participants believed they would be competing one-on-one in a trivia competition involving an "easy" category (e.g., current events, TV sitcoms), they were severely overoptimistic about winning because they primarily considered how well they could do on such a category, not how well their competitor could do (Windschitl et al., 2003). However, for a "difficult" category (e.g., European politics, home insurance facts), they were severely underoptimistic because they considered how poorly they would do on the category more so than how poorly their competitor would do. This pattern of results in which people are overoptimistic when circumstances are generally easy and underoptimistic when they are generally hard has been called the shared-circumstance effect (SCE; Windschitl et al., 2003).

Overoptimism and underoptimism can have important consequences. Overoptimism can lead to

* Corresponding author. Fax: +1 319 335 0191.
E-mail address: jason-rose@uiowa.edu (J.P. Rose).

0749-5978/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.obhdp.2007.08.003

misallocated time, effort, or resources into a competitive endeavor (e.g., starting a business, becoming a professional athlete, waging a war) and could eventually promote various adverse effects, including psychological effects, such as disappointment, regret, frustration, or depression (Burson, Larrick, & Klayman, 2006; Camerer & Lovallo, 1999; Dunning, Heath, & Suls, 2004; Moore & Kim, 2003). Underoptimism can lead to the avoidance of a potentially beneficial situation, such as when females, who were equally qualified to male counterparts in science abilities, expressed less interest in entering a lucrative science contest (Ehrlinger & Dunning, 2003).

Given the ubiquity of competition in everyday life—involving activities such as sports, politics, educational scholarships, war, employment—and given the potential negative effects of overoptimism or underoptimism caused by egocentric thinking, it is important to attempt to understand when and how people might avoid thinking egocentrically when in competitive contexts. In the present work, we investigated whether people can learn their own way out of egocentrism through their repeated experiences within a given type of competition. In all but one of the published studies of shared-circumstance effects (SCEs), people have made likelihood judgments about "single-play" challenges (the exception is a recently published study by Moore and Cain (2007), which is discussed later). In these "single-play" studies (e.g., Moore & Kim, 2003; Windschitl et al., 2003), participants have made likelihood judgments about the outcomes of competitions in which they are (or will be) involved. Researchers have not investigated whether participants' experiences with a previous competition—including knowledge of its outcome—will affect participants' next set of predictions. Given that people often participate in "repeated-play" challenges (e.g., athletes and armies often compete against the same competitor on multiple occasions and in many college courses each exam is a competition for top scores), we thought it was important to investigate the influence of repeated plays on the tendency for participants to think egocentrically and to exhibit the overoptimism and underoptimism that are characteristic of SCEs.

Before describing our specific paradigm and predictions, we will first discuss the following in the next two sections: (1) the causes of egocentrism and SCEs and (2) previous research that is generally relevant to the issue of how people might or might not effectively learn from repeated-play contexts.

Why are people egocentric?

When judging the likelihood of winning a competition, people must consider not only how their strengths, weaknesses, and the general circumstances bode for their performance but also how their competitors' strengths, weaknesses, and the general circumstances bode for their respective performances. We use the term *egocentrism* in a general way to refer to the idea that when people are asked about their optimism about winning, self-relevant assessments (e.g., thoughts/projections about their own performance) are weighted more heavily than competitor-relevant assessments (Kruger, 1999; Ross & Sicoly, 1979; Windschitl et al., 2003). There are, in fact, numerous types of egocentrism accounts and other accounts (e.g., focalism) that articulate reasons why there may be differential weighting of self-relevant and competitor-relevant assessments under various conditions (see reviews by Chambers & Windschitl, 2004; Moore, 2007). We will next discuss a subset of accounts that are most relevant to the present work.

To begin, it is helpful to distinguish accounts that suggest *irrational* reasons for differential weighting from accounts that suggest *rational* reasons. We use the term *irrational* to refer to processing biases that would hurt rather than help judgment accuracy in the relevant context and that are not apparently grounded within a broader, adaptive processing strategy. Among the accounts citing irrational processes, one type (or subtype) of egocentrism account suggests that people's tendency to place greater weight on self-relevant information is simply due to a *chronic attention bias*, whereby people typically think more about the self than about others (see e.g., Chambers, Windschitl, & Suls, 2003; Eiser, Pahl, & Prins, 2001; Ross & Sicoly, 1979; Windschitl et al., 2003). Examples of other irrational accounts for differential weighting include ones that concern egocentric anchoring-and-insufficient-adjustment, a generic focalism bias, or general difficulties in considering evidence about groups of competitors rather than single competitors (for a review, see Chambers & Windschitl, 2004).

The accounts that posit *rational* reasons for differential weighting suggest that self-relevant assessments carry more weight than do other-relevant assessments because we often have more information or better information about the self than others (see e.g., Ross & Sicoly, 1979). For example, a differential-confidence account described by Chambers and Windschitl (2004) suggests that although people may believe that both the self and a competitor have low skill at a hard task (or high skill at an easy task), they are more confident regarding their self-assessments (see also Kruger, Windschitl, Burrus, Fessel, & Chambers, in press; Moore & Small, 2007). Hence, when gauging their optimism about winning at the task, they give more weight to the self-assessment—yielding underoptimism on hard tasks and overoptimism on easy tasks.[1]

In general, both irrational and rational sources of differential weighting have been shown to contribute to SCEs (e.g., Krizan & Windschitl, 2007a; Kruger et al., in press; Moore & Kim, 2003; Moore & Small, 2007; Windschitl et al., 2003). As we will discuss later, the fact that both rational and irrational sources of differential weighting can fuel SCEs has important implications for the manner and ease by which people might learn their way out of differential weighting.

The influence of repetition, feedback, and experience on judgment accuracy

Repetition, feedback, and experience tend to have a positive influence on mitigating judgmental biases (e.g., Cox & Grether, 1996; Einhorn & Hogarth, 1978; Garb, 1989; Kagel & Levin, 1986; Larrick, 2004; Thiede & Dunlosky, 1999; Van Boven, Loewenstein, & Dunning, 2003). For instance, Kagel and Levin (1986) showed that participants overbid in early rounds of an auction (i.e., participants exhibited a winner's curse; Bazerman & Samuelson, 1983), but placed bids more optimally as they gained experience (although see Ball, Bazerman, & Carroll, 1991). In a study by Van Boven and colleagues (2003), participants showed lower levels of the endowment effect (i.e., ascribing greater value for an owned object) as they gained experience across repeated trials. Also, in a study by Cox and Grether (1996), preference reversals between choice and valuation became smaller across repeated trials. Finally, providing either performance feedback or calibration feedback about past trials has been shown to reduce over-confidence (i.e., improve calibration) on future trials (Arkes, Christensen, Lai, & Blumer, 1987; Radhakrishan, Arrow, & Sniezek, 1996; Renner & Renner, 2001; Schraw, Potenza, & Nebelsick-Gullet, 1993; Zakay, 1992; Zechmeister, Rusch, & Markell, 1986; although see Fischer, 1982; and Koriat, Sheffer, & Ma'ayan, 2002).

Although repetition, experience, and feedback seem to improve judgment biases, there are also situations in which people miss opportunities to apply insights from one task to another. For instance, people often fail to recognize the applicability of their experience gained from past contexts or trials (where learning occurred) to future contexts or trials (where learning must be applied). Indeed, research on analogical transfer in problem solving suggests that problem solvers have difficulty transferring what they have learned from past problems to future problems that share structural similarities (e.g., both problems involve a division of elements), but not surface similarities (e.g., one problem is about apples and the other is about automobiles) (Bassok, 2003; Holyoak & Koh, 1987; McClosky, 1983; Novick, 1988; Reed, 1987; Reed, Dempster, & Ettinger, 1985). Similarly, in Van Boven and colleagues (2003), participants eventually showed a reduction in the endowment effect for the *same* commodity across five rounds, but the effect rebounded when a *novel* commodity was introduced. Finally, Kagel and Levin (1986) showed that, despite becoming more optimal in their bidding across rounds in 3-person auctions, participants again provided sub-optimal bids upon switching to 6-person auctions. In sum, repetition, feedback and experience often have very specific (i.e., non-generalized) influences on performances and judgments, perhaps because of the difficulties that people have in recognizing deep, structural similarities across contexts or trials (Bassok, 2003).

To some degree, the central issue being investigated in the present paper—whether and how people might learn their way out of egocentrism in a repeated-play situation—is akin to a problem-solving issue in which analogical transfer can play a critical role. We assumed that people's ability to avoid egocentrism and make accurate predictions depends on the extent to which they perceive that their predictions and the feedback on previous plays are relevant to their predictions about the impending play. Hence, as we explain in more detail later, the issues of surface and structural similarities have relevance to the "problem-solving performance" of participants in our experiments.

The paradigm used in the present studies

Each study in this project involved a modified version of the trivia-competition paradigm used in studies by Windschitl et al. (2003). As mentioned earlier, participants in those original studies estimated the likelihood of victory over a co-participant in each of many trivia categories—some of which sounded very difficult (e.g., European politics) and some of which sounded very easy (e.g., current events). Participants were substantially overoptimistic for a victory in easy categories and underoptimistic in difficult categories, even though the

[1] It is important to note that a *differential-regression account* provides a second explanation as to why having more knowledge about the self than others can ultimately lead to SCEs (Chambers & Windschitl, 2004; Moore, 2007). Namely, people may simply make less extreme assessments about others' skills (for either hard or easy tasks) than they do about their own skills. This would thereby make people pessimistic for hard tasks and optimistic for easy ones, even if people did not give more weight to the self-assessments than to other-assessments. Recent research comparing the differential-regression account to the differential-confidence account suggests that both accounts (along with others) can help explain portions of SCEs (Kruger et al., in press; Moore & Small, 2007). In this work, we focus primarily on accounts that assume differential weighting. As described later, this focus is generally consistent with our results, although these results do not preclude a role for differential regression.

average objective likelihood of victory in the one-on-one competitions was necessarily 50%.

The main modification that we introduced to this paradigm was that participants played in multiple rounds of trivia. The number of rounds and the specific elements of those rounds differed across the four studies in this project. Below we outline the specific elements of Study 1, which will provide some context for our subsequent discussion of the types of processes that could result in a reduction in SCEs across rounds of play.

In Study 1, there were 3 rounds. At the beginning of the first round, participants estimated the likelihood of winning in each of 10 possible categories (5 hard and 5 easy), and also provided estimates of the perceived level of knowledge for both the self and their competitor in each of the 10 categories. They then played a subset of these categories (i.e., the participant and co-participant separately completed the same quiz questions). Finally, participants either received feedback about how the self and another competitor scored (full-feedback condition), or only received information about their own performance (self-only feedback condition). This procedure was then repeated two additional times, with each round involving a new set of 10 categories. We should emphasize that the feedback was always completely valid, and tie-breakers were used to ensure there was exactly one winner per category.

Hypotheses regarding repeated plays and SCEs

We expected to detect robust SCEs in the initial round of all our studies, with participants being overoptimistic about victory in the easy categories but underoptimistic in the hard categories (Windschitl et al., 2003). However, the critical issue was whether and how participants would modify their levels of optimism in subsequent rounds—and how this would differ as a function of the type of feedback they received. We did not expect main-effect shifts in optimism. Instead we were focused on whether the SCE would shrink and/or disappear with the help of repeated plays.

Why might the overoptimism and underoptimism embodied by SCEs shrink because of repeated plays? Consider the situation of the participants in the full-feedback condition of Study 1, who learned not only how well/poorly they did on categories but also how well/poorly their competitor did. Most participants in this condition learned that when they did well on a category, their competitor did also; and when they did poorly, their competitor did too. It seemed possible to us that this information would lead to more optimistic predictions about winning in hard categories and less optimistic predictions about winning in easy categories. More specifically, the salience of the fact that the two scores were both bad/good and that the determination of the winner depends on *both* scores may be enough to reduce irrational forms of egocentrism (or focalism).[2]

Again, we used natural feedback rather than uniformly disconfirming feedback, so one participant always learned that they did at least slightly better or worse than the other. Hence, it would be unrealistic for us to expect that participants, on the whole, would *quickly* learn to completely avoid egocentrism. In fact, if participants were only attentive to whether they won or lost (not the closeness of scores), they might learn to avoid egocentrism rather slowly because the win/loss feedback would contradict people's predictions only on 50% of occasions. They might learn faster if they attend to the fact that their competitor's score was generally similar to theirs (high on easy categories and low on hard ones). Regardless of whether the learning would be quick (evident by Round 2) or somewhat slower, we were optimistic that some sensitivity to egocentrism would be developed and that when people made their likelihood estimates, they would do more to consider the expected performance of their competitor for each category, not just their own expected performance. We presumed that the ability of participants to transfer insight from a previous round (e.g., "I can still win a category if my competitor does poorly too") to the way in which they made predictions about subsequent rounds would be facilitated by the fact that there was high surface similarity between the two rounds (both involved the same game, the same prediction task, and the same competitor).

Short of this high level of insight that participants might have during the repeated plays, there are two other types of reasons why people might exhibit a reduction in SCEs. First, the extent to which a participant saw his/her competitor as a concrete and familiar person may have increased from round to round, especially when the participant received feedback about that person's performance. Previous research on above-average effects has shown that such effects are reduced when a comparison referent is concrete. That is, when a participant compares him or herself to an individuated referent (versus an ambiguous referent, such as the average college student), they are less susceptible to reporting above-average perceptions of their traits and performances (Alicke, Klotz, Breitenbecher, Yurak, & Vrendenburg, 1995; Buckingham & Alicke, 2002). Also, SCEs are less severe when the participant knows his/her competitor well rather than poorly (Windschitl et al., 2003). Hence, shared-circumstance effects in our Study 1 might be reduced across

[2] Henceforth, we will not distinguish focalism from egocentrism. In our experiments, the self was always the target, which makes focalism indistinguishable from an egocentric-attention account. Yet, it is important to acknowledge that effects that have been attributed to egocentrism might also or instead be attributable to focalism.

rounds as a competitor becomes more concrete and familiar. Second, participants may simply begin to recognize that their probability estimates were not very accurate—that they were sometimes winning categories about which they expressed very low optimism and losing categories about which they expressed high optimism. This may make them reluctant to make extreme predictions in either direction and instead to regress their probability judgments toward 50%, which would reduce SCEs.

In sum, there are at least three types of reasons why SCEs might be reduced across rounds in the first study (and subsequent studies): (1) insights relevant to the egocentric bias, (2) the enhanced familiarity/concreteness of the competitor, and (3) greater reluctance to make extreme probability estimates. These reasons are not mutually exclusive.

In each of our studies, we tested for a reduction in SCEs at the group level. Our concern was whether people would tend to use naturally occurring feedback in a way that reduced the extent to which they exhibited the overoptimism/underoptimism of the SCE. We were less interested in whether specific individuals who received specific feedback would alter their optimism on future rounds. Indeed, change in optimism at the individual level is not readily interpretable in this study, because so many factors vary at one time for a given individual.

Study 1

The key elements of Study 1 have already been described. The study was an initial test of whether people would modify their optimism in a way that reduces the SCE across three rounds of a trivia competition. We expected reductions in the SCE to be greatest in the full-feedback condition in which participants learned their own score and the competitor's score for each category played in each round. This feedback offered an opportunity for participants to notice that both scores are critical and perhaps that the competitor's scores were roughly as bad (good) as their score on hard (easy) categories. These forms of insight could reduce egocentric thinking if people transfer their learning to subsequent rounds. Also, the extent to which the competitor is concrete and familiar could increase over rounds in both conditions, but more so in the full-feedback condition. Finally, participants could only notice that their predictions were overly optimistic or overly pessimistic in the full-feedback condition. In the self-only feedback condition, which is essentially a control condition, participants did not learn who won (or the scores of the other person) and therefore could not gauge the accuracy of their probability judgments.

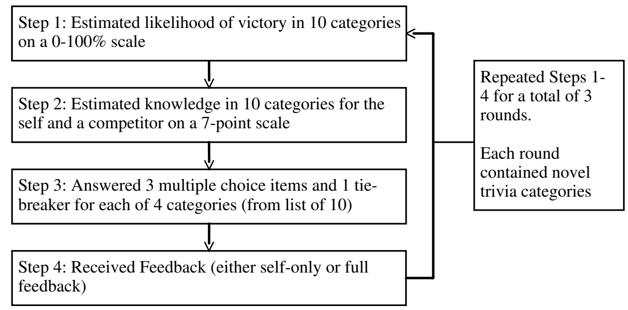

Fig. 1. Procedures in Study 1.

Method

Participants and design

Fifty-six students from the University of Iowa participated to fulfill a research-exposure component for their elementary psychology course. The design was a $2 \times 2 \times 3$ mixed design, where the type of feedback (full feedback or self-only feedback) was manipulated between participants and category difficulty (easy or hard) and round (1, 2, or 3) were manipulated within participants. Two counterbalancing manipulations were also part of the design and are described below in the procedure section.

Materials: Trivia categories and questions

Thirty trivia categories (taken from Windschitl et al., 2003) were split into 3 blocks of 10, with 5 easy and 5 hard per block. For each block, we prepared quizzes for 4 of the categories (2 easy and 2 hard categories). Appendix 1 displays the 10 categories for each block and the subset of categories for which quizzes were created. Each quiz contained 3 multiple-choice questions and 1 tie-breaker—all based on the category theme (see Appendix 2 for sample questions). The multiple-choice questions were designed and pilot tested to be easier for the easy categories and harder for the hard categories. (Analyses of quiz performances for this study and subsequent studies indicate we were always successful.) The tie breakers asked for responses on continuous scales (e.g., historic dates).

Procedure

Fig. 1 contains a flow chart representing the procedures for Study 1. Participants entered the sessions in pairs, but were generally unfamiliar with each other.[3] First, participants were informed that they would shortly be playing a round of trivia against the other person and were given a questionnaire that listed 10 categories. The instructions stated:

[3] Only 6% of our participants indicated they were friends. The results look identical when friends were included or excluded, thus we decided to include all participants in our analysis.

Four trivia domains will be randomly selected from the list below. You will be asked three questions in each of the selected domains. Whoever answers more questions correctly in a given domain will be the winner of that domain. In the event of a tie in a given domain, the answers you supply to tie-breaker questions will be used to decide the domain winner.

Next, participants estimated the likelihood that they would win in each of the 10 categories (*The chance you will win the Pop Culture category is ______ %*). Participants were instructed to answer with any value between "0" and "100", where a response of "0" would indicate no chance of a win, a response of "50" would indicate an equal chance of a win, and a response of "100" would indicate certainty for a win. Participants then made 10 knowledge estimates about the self and 10 about their competitor, with the order of these two blocks counterbalanced. More specifically, participants estimated their own and their competitor's knowledge for each category on a 7-point scale (1 = *very little knowledge*; 7 = *a great deal of knowledge*).

Next, participants privately completed the quizzes for 4 categories (2 easy and 2 hard). Participants in a given pair always completed the same quizzes. The experimenter scored the quizzes privately, and then gave the participants feedback. Approximately half of the dyads received full feedback about their performances, which was accurate feedback about who won each category and about the number of multiple-choice questions both competitors answered correctly in each of the categories. Full feedback was provided visually in front of both participants (i.e., written on a chalkboard) and stated orally by the experimenter. The other half of the dyads received self-only feedback, which was accurate feedback about how many questions the self answered correctly in each category. Self-only feedback was provided privately and no information about the competitor was given. Critically, after receiving feedback in Round 1, participants competed in two additional rounds of trivia, where each new round involved making likelihood and knowledge judgments about 10 new categories, completing quizzes for a subset of 4 new categories, and receiving feedback (see Fig. 1).

We counterbalanced the ordering of categories, such that a given category had an equal chance of appearing in Round 1 as in Round 2 or Round 3. This counterbalancing factor yielded no significant interactions with the main dependent measures and will not be discussed further.

Results

Likelihood judgments

For each participant and within each round, we calculated the composite values (i.e., means) of his/her probability estimates for the 5 hard categories and the 5 easy categories. Further data analyses were conducted on these composite values. Table 1 lists the means of these composites by round and by feedback manipulation. Table 1 also displays the mean SCEs (i.e., the difference between probability estimates for easy and hard categories) by round and by feedback manipulation.

Before assessing whether the SCE changed across rounds as a function of feedback type, we first confirmed that the effect actually existed in the first place—i.e., in Round 1. We compared Round 1 mean probability estimates from the easy and hard categories with 50%, which is the objective average probability of winning (each category always had exactly one winner and one loser). Indeed, participants were significantly overoptimistic about victory in easy categories ($M = 57.8$, $SD = 14.3$), $t(55) = 4.09$, $p < .01$, and underoptimistic about victory in hard ones ($M = 25.6$, $SD = 15.7$), $t(55) = 11.67$, $p < .01$.

Table 1
Mean Likelihood and Absolute Knowledge Judgments in Study 1 by Category Difficulty, Round and Type of Feedback ($N = 56$)

	Full Feedback			Self-only Feedback		
	Round 1	Round 2	Round 3	Round 1	Round 2	Round 3
Likelihood Judgments						
Easy						
M	64.7*	63.5*	62.8*	56.1*	57.6*	58.7*
SD	13.3	17.9	15.3	13.4	16.2	15.3
Hard						
M	29.5*	30.3*	29.0*	23.6*	27.1*	28.5*
SD	17.1	11.9	13.7	14.5	15.2	16.3
SCE						
M	35.7*	33.1*	33.7*	32.5*	30.5*	30.1*
SD	16.8	16.8	17.4	13.2	16.9	18.3
Absolute knowledge judgments						
Easy–Self						
M	5.18	5.28	5.22	4.97	5.14	5.16
SD	0.68	0.99	0.84	0.77	0.90	0.93
Easy–Other						
M	5.31	5.32	5.06	5.03	5.11	5.20
SD	0.53	0.62	0.76	0.95	0.79	0.82
Hard–Self						
M	2.34	2.35	2.46	2.24	2.31	2.44
SD	0.93	0.67	0.81	0.83	0.82	0.87
Hard–Other						
M	2.45	2.48	2.42	2.78	2.62	2.63
SD	0.82	0.91	0.69	0.85	0.83	0.72

Note. Likelihood judgments were made on a 0–100% scale. An asterisk in the *Hard* or *Easy* rows under *Likelihood Judgments* indicates that the mean was significantly different from 50% ($p < .05$). The shared-circumstance effect (SCE) was calculated by subtracting the mean likelihood estimate for hard categories from the mean for easy categories. An asterisk in the *SCE* row indicates that the SCE was significantly different from 0 ($p < .05$). Absolute knowledge judgments—reported in separate rows for the self and other—were made on a 1–7 scale. All absolute judgment values are significantly different from 4, which was the scale's midpoint ($p < .05$).

Now we turn to the main question of whether these biases changed as a function of round and type of feedback. We submitted the composite probability estimates to a 2 (Type of Feedback) × 2 (Category Difficulty) × 3 (Round) ANOVA, with the last two factors being within subjects. The ANOVA detected a main effect for category difficulty, $F(1,54) = 313.12$, $p < .01$, where easy categories elicited more optimism for a victory ($M = 60.2$) than hard categories ($M = 27.8$). No other significant differences were detected (all $ps > .10$). Most importantly, the three-way interaction was not significant, suggesting that participants were not utilizing comparative feedback (available in the full-feedback condition only) in a manner that produced an attenuation of the SCE. As can be seen in Table 1, the SCE in the full-feedback condition was a mere 2% less in Round 3 (33.7) than in Round 1 (35.7).

In more detailed analyses, we tested whether any significant linear trends (increases or decreases in optimism) could be detected across rounds for each of the four combinations of category difficulty and type of feedback. This was explored—separately for each combination—by regressing probability judgments onto the specific round (1, 2, or 3) that judgments were elicited (Kleinbaum, Kupper, Muller, & Nizam, 1998). None of these four analyses revealed a significant linear or quadratic trend (all $\beta s < |.15|$, $ps > .10$). In other words, there was no significant evidence that participants' optimism changed across rounds in either of the feedback conditions.

Knowledge estimates

For each participant and within each round, we calculated the means of his/her estimates of knowledge for the 5 hard and 5 easy categories. The same was done for his/her estimates of the competitor's knowledge in the categories. These means were submitted to a 2 (Type of Feedback) × 2 (Target: Self or Other) × 2 (Category Difficulty) × 3 (Round) ANOVA, with the last three factors being within subjects. Not surprisingly, the category difficulty main effect was robust, $F(1,55) = 675.08$, $p < .01$, with participants reporting higher knowledge estimates for easy ($M = 5.16$, $SD = .63$) than for hard categories ($M = 2.64$, $SD = .62$). A Target × Round interaction was also significant, but this effect was relatively weak and not very meaningful for the issues at hand, $F(2,53) = 3.69$, $p < .05$. Most important is the fact that no other significant main effects or interactions emerged (all $ps > .05$).

Table 1 displays the mean knowledge estimates by target, category difficulty, round, and type of feedback. As the means suggest, participants viewed the self and a competitor as having roughly equal (and good) knowledge of the easy categories overall, $t(55) = -.16$, $p > .10$. Participants also viewed the self and a competitor as having roughly equal (and poor) knowledge of the hard categories overall, although a t-test did detect a small but significant difference between knowledge ratings for the self and the competitor on hard categories, $t(55) = -2.64$, $p < .05$. As will be discussed below, the results of these two t-tests are important when juxtaposed against the fact that participants showed robust overoptimism in their probability estimates about winning easy categories, and robust underoptimism about winning hard categories.

Weighting of self-knowledge assessments versus competitor-knowledge assessments

Was differential weighting of self- and other-knowledge assessments a factor in causing the SCEs? A commonly used technique in the literature for detecting differential weighting in comparative judgments is to use a path-analytic approach (e.g., Klar & Giladi, 1997; Kruger, 1999; Moore & Kim, 2003; Windschitl et al., 2003). From the data, we could determine what better predicts participants' levels of optimism, their estimates of self-knowledge or their estimates of their competitor's knowledge. Thus, for each participant and within each round, we conducted a path analysis treating his/her probability estimates as the dependent variable and his/her self-knowledge and competitor-knowledge ratings as predictor variables. See Table 2 for the results from these path analyses, with separate values reported for each round and type of feedback. If participants were initially egocentric, then their self-knowledge estimates would be better predictors of optimism in Round 1 than would their competitor-knowledge estimates. This was indeed the case in both feedback conditions. For example, in the full-feedback condition in Round 1, the mean beta for self-knowledge estimates was .74 whereas the mean beta for competitor-knowledge estimates was .09. The absolute value of the former was significantly greater than the absolute value of the latter, $t(19) = 2.97$, $p < .01$. This pattern was remarkably similar within all rounds and in both feedback conditions, suggesting that differential weighting was robust at Round 1 and did not fade.

We should note that interpretations of the results from this type of path-analytic approach can be problematic if participants' knowledge estimates about others are significantly more regressive than their estimates of themselves. This is a particularly likely scenario for studies in which participants are asked to rate themselves and the *average* peer and when the researcher conducts the path analysis across participants and within a given skill/category. In those cases, there is bound to be relatively little variability (more regressiveness) in knowledge judgments about the average peer. Thus, when judgments of the self and the average peer are entered as predictor variables in a regression/path analysis, the former will naturally explain more variability in comparative or probability judgments (see Burson

& Klayman, 2006; Moore, 2007). However, for our path analyses, participants were asked to rate a *specific* peer on *multiple* categories, and the regression/path analysis was conducted across those ratings. Although researchers might expect that participants would be regressive in how they rate the category knowledge of a peer they just met (i.e., less extreme in knowledge ratings for hard and easy categories), this differential regressiveness was, in fact, relatively small. As we reported earlier, participants rated the self and a competitor as having roughly equal (and good) knowledge of the easy categories, and there was only a small difference in how they rated the self and a competitor for the hard categories. Furthermore, when we restricted our path analyses to participants for whom the difference in knowledge ratings between easy and hard categories was greater for the ratings of their competitor than for the self ($n = 22$), the resulting betas continued to suggest egocentric weighting (e.g., across all categories, $M_{\text{self beta}} = 0.97$ and $M_{\text{other beta}} = -0.08$). Similar findings were also found when the restriction was based on having greater overall variability (standard deviations) across competitor-knowledge judgments than self-knowledge judgments.[4] Therefore, in the context of the present study, we suggest that these path analyses do reflect differential weighting, and can also be used to track the reduction (or non-reduction) in differential weighting across rounds.

Discussion

Consistent with expectations, likelihood estimates about a victory initially (in Round 1) elicited very robust shared-circumstance effects. That is, participants tended to report a high likelihood about winning in easy categories and a low likelihood about winning in hard categories. Interestingly, participants showed these extreme levels of over- and under-optimism despite viewing both the self and a competitor as similarly knowledgeable in both easy and hard categories. Additionally, path analyses showed that judgments of likelihood were much more influenced by self-knowledge assessments than competitor-knowledge assessments, suggesting the presence of egocentric processing. Did experience with the competitive context, the same judgment task, and a competitor influence egocentrism or SCEs across rounds? The answer is generally no. Across rounds two and three, participants in both types of feedback conditions continued to be egocentric, and the magnitude of the SCEs did not diminish across rounds. Recall that in the introduction, we had mentioned three reasons why SCEs might be reduced across rounds in this study: (1) insights relevant to the egocentric bias, (2) enhanced familiarity/concreteness of the competitor, and (3) greater reluctance to make extreme probability estimates. Apparently, none of these factors had a significant impact on participants' responses in Study 1.

Why were egocentrism and SCEs so persistent? We will describe 4 interrelated possibilities. First, the fact that we required participants to make a large number of likelihood judgments per round (10) may have made it difficult for participants to diligently apply insights from a previous round when making new judgments. In other words, knowing that they had to give 10 judgments might have caused participants to find simple ways of responding rather than applying feedback.

Second, although there were numerous layers of surface similarity between the tasks in different rounds (i.e., participants made the *same* types of judgments, in the *same* competition, against the *same* competitor), a key dissimilarity might still be a critical factor in precluding participants from using insights prompted by feedback to shape optimism. Namely, participants might see the feedback about a previous category as relevant only to optimism about that same category, not to optimism about *different* categories. As mentioned earlier, the transfer-of-learning literature suggests that people have difficulty applying insights from one task to another when only deep structure is shared by the two tasks—i.e., no obvious surface similarity to facilitate the recognition of the deep-structure similarity (Bassok, 2003).

A third possible explanation as to why SCEs were persistent was that participants perceived the quality and/or quantity of the full feedback as inadequate for altering their optimism. The quiz for each of the categories about which they received feedback had only 3 questions and 1 tie-breaker. Perhaps participants felt their performance and/or that of their co-participant were greatly susceptible to random factors and thus uninformative. In essence, they may have viewed feedback as a curiosity, but irrelevant to their future optimism.

A fourth and different type of explanation is that, within this paradigm, rational forms of egocentrism (e.g., differential confidence) are a strong source of differential weighting. Although previous research has already shown that irrational focalism or differential attention can contribute to differential weighting in this type of paradigm (e.g., Study 4 of Windschitl et al., 2003), perhaps the independent contribution of irrational biases to differential weighting is small in this paradigm relative to what is caused by a rational weighting

[4] As an alternative way of examining the differential weighting issue, we conducted hierarchical regression analyses in which a participant's likelihood judgments were first regressed on self–other difference scores (i.e., self-knowledge estimate minus other-knowledge estimate), and then self-knowledge estimates were added as a predictor variable in a second step (see Burson & Klayman, 2006). In these analyses, the self-knowledge estimates that were added in the second step tended to account for significant variance ($p < .01$)—beyond what was accounted for by the self–other difference scores. The overall average R-squared change was 19% (and this value did not substantially shift as a function of round). This finding is consistent with our conclusion that egocentric weighting influenced likelihood judgments.

Table 2
Results for path analyses from Study 1 ($N = 56$)

	Full Feedback				Self-only Feedback			
	β (Self)	β (Other)	Differential	r	β (Self)	β (Other)	Differential	r
Round 1	.74	.09	.83	.79	.81	.05	.86	.72
Round 2	.77	.10	.87	.80	.79	.08	.87	.79
Round 3	.84	.02	.86	.83	.76	.07	.83	.77

Note. For each participant, a path/regression analysis was conducted in which his/her likelihood estimates were regressed on his/her ratings of knowledge for the self and for the competitor ("other"). The values in the β (*Self*) and β (*Other*) columns are the mean betas for self and other estimates from these regression analyses. The differential column conceptually reflects the extent to which the self and other assessments have equivalent weight in predicting likelihood judgments. For example, a differential of zero would reflect that the betas for self and other assessments were equal but in opposite directions (+/−). In cases where β (*Other*) is negative, the differential is β (*Self*) minus the absolute value of β (*Other*). In cases where β (*Other*) is positive, which is opposite of the normative direction, the differential is β (*Self*) plus β (*Other*). The values in the r column reflect the mean correlation between self- and other-knowledge estimates.

strategy (e.g., differential confidence). If this rational weighting strategy was a strong factor in Round 1 of Study 1, then it would naturally have been difficult to prompt participants to give less weight to self-assessments in Rounds 2 or 3; participants may have used a sensible strategy of basing their optimism on what they knew about themselves more than assumptions about their competitor.

Overall, any or all of these four interrelated reasons could be responsible for preventing our participants from learning to avoid SCEs. These reasons motivated methodological changes for Study 2.

Study 2

Study 2 resembled Study 1 both conceptually and operationally, except that we introduced changes that we believed might make it easier for participants to learn to avoid SCEs across rounds. First, to address the issue of participants being overwhelmed with a large amount of judgments, we decreased the amount of categories judged in a given round from 10 to 4. Second, to address the issue of whether or not the quantity and/or quality of feedback were meaningful enough to prompt the utilization of feedback, we increased the number of multiple-choice questions per category from 3 to 6. Third, to increase the perception of surface similarity across rounds, participants made judgments about the same four trivia categories (2 easy, 2 hard) across all rounds. Participants were repeatedly quizzed and received feedback on 2 of these categories (1 easy and 1 hard; hereafter called the *quizzed categories*). The other 2 categories were left *non-quizzed*. We believed that repeating the quizzed categories would—relative to the conditions of Study 1—increase the extent to which participants would view feedback and experience from a previous round as meaningful for future rounds. This would also presumably give people a greater sense of confidence about what their competitor did or did not know, thereby making a rational weighting strategy more balanced (i.e., less differential weighting for judgments about the quizzed categories in Rounds 2 and 3). Furthermore, we thought that including the repeatedly judged but non-quizzed categories would give us an opportunity to test whether reductions in egocentrism and the SCE would be specific to only the quizzed categories or if learning could be transferred to other, non-quizzed categories.

Method

Study 2 was identical to Study 1, with the following exceptions. Participants ($N = 44$) provided likelihood and knowledge estimates regarding four rather than ten categories per round. These four categories were the same in Rounds 1, 2, and 3. Two of the categories were easy (Pop Culture and Fast Food Chains) and two were hard (50's Movies and Baroque Music). However, each pair of participants only competed in one easy category and one hard category per round (i.e., they took quizzes for these categories and received feedback). For half of the participant pairs, the quizzed categories were always Pop Culture and 50's Movies, whereas for other pairs, the quizzed categories were always Fast Food Chains and Baroque Music. This counterbalancing did not influence the results and will not be discussed further. An additional difference from Study 1 was that the quiz for a category contained 6 questions (rather than 3) and a tie-breaker. Finally, although Study 2 included both a full-feedback condition and a self-only feedback condition, we report the results for only the full-feedback condition.[5]

[5] The results from the self-only feedback condition held no surprises and can be summarized as follows: (1) The results essentially replicated those from the same condition of Study 1, (2) SCEs were robust across rounds, and (3) Path analyses revealed an overweighting of self information that did not change across rounds. Omitting a full description of these results greatly simplifies the Results section and figures for Study 2 without losing critical information.

The design was a $2 \times 2 \times 3$ factorial, in which category difficulty (easy or hard), quiz type (quizzed or non-quizzed) and round (1, 2, or 3) were all manipulated within subjects.

Results

Likelihood judgments

Mean likelihood judgments—as a function of category difficulty, round, and quiz type—are listed in Table 3. As in Study 1, preliminary analyses confirmed that a SCE existed in Round 1, $t(43) = -5.5$, $p < .01$. Next, we submitted participants' likelihood estimates to a 2 (Category Difficulty) × 3 (Round) × 2 (Quizzed vs. Non-quizzed) ANOVA. This ANOVA detected a robust main effect for category difficulty, $F(1,43) = 52.7$, $p < .01$, where the easy categories elicited higher likelihood of victory estimates ($M = 60.7$, $SD = 14.9$) than hard categories ($M = 36.3$, $SD = 16.6$). Additionally, a significant Round × Category Difficulty interaction, $F(2,41) = 3.40$, $p < .05$, revealed that estimates for hard categories increased slightly across rounds (M in Round $1 = 34.2$; M in Round $3 = 39.0$), whereas likelihood estimates for easy categories stayed consistent across rounds (M in Round $1 = 60.5$; M in Round $3 = 60.9$). Finally, the Round × Category Difficulty × Quiz Type interaction was significant, $F(2,41) = 3.8$, $p < .04$. Given this three-way interaction and our a priori interests in specific patterns of means, follow-up tests were conducted.

Table 3
Mean Likelihood and Absolute Knowledge Judgments in Study 2 by Category Difficulty, Round, and whether a Category was Quizzed ($N = 44$)

	Quizzed Categories			Non-quizzed Categories		
	Round 1	Round 2	Round 3	Round 1	Round 2	Round 3
Likelihood Judgments						
Easy						
M	60.7*	65.6*	62.2*	60.3*	57.7*	58.1*
SD	18.5	20.6	21.3	18.0	19.9	19.8
Hard						
M	35.0*	34.6*	41.4*	34.0*	37.5*	37.2*
SD	19.7	17.1	23.2	23.9	22.2	23.2
SCE						
M	25.7*	31.0*	20.8*	26.3*	20.2*	21.0*
SD	23.5	31.0	29.3	30.2	30.4	32.9
Absolute Knowledge Judgments						
Easy–Self						
M	4.91	5.25	5.20	5.20	5.02	4.98
SD	1.15	1.24	1.24	1.32	1.37	1.45
Easy–Other						
M	5.34	5.18	5.22	5.34	5.18	5.11
SD	1.05	1.16	1.26	1.05	0.99	0.97
Hard–Self						
M	2.77	2.34	2.36	2.81	2.88	2.84
SD	1.32	1.18	1.16	1.38	1.45	1.48
Hard–Other						
M	3.15	3.09	3.16	3.14	3.11	3.02
SD	1.09	1.34	1.22	1.32	1.08	1.10

Note. An asterisk in the *Hard* or *Easy* rows under *Likelihood Judgment* indicates that the mean was significantly different from 50% ($p < .05$). An asterisk in the *SCE* row indicates that the SCE was significantly different from 0 ($p < .05$). All absolute judgment values are significantly different from the midpoint (4) ($p < .05$).

The most critical finding from these follow-up tests was that the Category Difficulty × Round simple interaction was not significant within the non-quizzed condition ($p > .1$), but it was significant in the quizzed condition, $F(2,41) = 3.3$, $p < .05$. Paired-samples t-tests confirmed that the SCE within the quizzed condition was slightly smaller in Round 3 ($M = 20.8$) than in Round 1 ($M = 25.7$) or 2 ($M = 31.0$), $t(43) = 2.02$, $p < .05$ and $t(43) = 2.53$, $p < .02$ (see Table 3). In short, it appears that for quizzed categories (but not non-quizzed categories) there was a slight reduction in the SCE.

Similar to Study 1, we also tested whether significant linear trends (increases or decreases in optimism) could be detected across rounds for each cell defined by the four combinations of category difficulty and quiz type. No significant linear (or quadratic) trends were detected (all βs $< |.15|$, ps $> .1$). The set of means closest to a significant linear trend was the one for quizzed/hard categories ($\beta = .12$, $p = .17$). A paired t-test indicated that likelihood estimates for a victory in hard categories increased slightly from Round 2 to Round 3, $t(43) = -2.19$, $p < .04$. Overall, it appears that participants' overoptimism about easy categories and underoptimism about hard categories changed little across rounds, with the possible exception of a slight mitigation of underoptimism between Rounds 2 and 3 for hard, quizzed categories.

Knowledge estimates

Table 3 also displays the mean knowledge ratings. Knowledge ratings were submitted to a 2 (Target: Self or Other) × 2 (Category Difficulty) × 3 (Round) × 2 (Quizzed or Non-Quizzed) ANOVA. Not surprisingly, a main effect for category difficulty was detected, $F(1,43) = 185.6$, $p < .01$. A main effect of target was also detected, $F(1,43) = 8.4$, $p < .01$, where knowledge estimates for the self were lower than estimates for a competitor. The Target × Category Difficulty × Round interaction was significant, $F(2,42) = 5.3$, $p < .01$, as was the 4-factor interaction, $F(2,42) = 8.5$, $p < .01$. An inspection of Table 3 reveals subtle patterns that comprise these interactions, but none that we find theoretically meaningful. What is most striking is how similarly the self and a competitor are perceived—with only subtle deviations from this pattern. In general, participants believed both the self and a competitor were highly knowledgeable in easy categories and not in hard ones.

Table 4
Results for path analyses from Study 2 ($N = 44$)

	Quizzed Categories				Non-quizzed Categories			
	β (Self)	β (Other)	Differential	r	β (Self)	β (Other)	Differential	r
Round 1	.80	−.05	.75	.55	.69	−.05	.64	.18
Round 2	.68	−.25	.43	.29	.86	−.25	.61	.59
Round 3	.53	−.07	.46	.30	.89	−.24	.65	.67

Note. The path analyses for Study 2 were conducted within category and across participants. The results were then averaged across categories. The differential column conceptually reflects the extent to which the self and other assessments have equivalent weight in predicting likelihood judgments. For example, a differential of zero would reflect that the betas for self and other assessments were equal but in opposite directions (+/−). In cases where β (*Other*) is negative, the differential is β (*Self*) minus the absolute value of β (*Other*). In cases where β (*Other*) is positive, which is opposite of the normative direction, the differential is β (*Self*) plus β (*Other*). The values in the r column reflect the mean correlation between self- and other-knowledge estimates.

Weighting of self-knowledge assessments versus competitor-knowledge assessments

Because participants made only 4 likelihood judgments per round, we could not compute separate path/regression analyses per participant, per round (as was done in Study 1). Instead, regression analyses were conducted separately for each category within a round, and the resulting betas were then averaged. Table 4 displays the results. Differential weighting was again very strong across all rounds for non-quizzed categories, replicating the findings in Study 1. Although this pattern is also strong for the quizzed categories, there is some preliminary evidence that participants began to weight the self and a competitor more equivalently by Rounds 2 and 3 when formulating optimism (see changes in "differential" column in Table 4).[6]

Discussion

Recall that for Study 2—relative to Study 1—we made methodological changes aimed at removing impediments that might keep people from learning to avoid egocentrism and SCEs. Similar to Study 1, egocentrism and SCEs were robust in Round 1. Despite our methodological changes, however, people continued to display robust egocentrism and SCEs across Rounds 2 and 3. There was some evidence that SCEs reduced between Rounds 2 and 3 for repeatedly quizzed categories, but this evidence was quite modest (albeit significant) and requires replication.

Furthermore, it is important to note that even though participants showed slightly reduced SCEs for the quizzed categories, there was no reduction in the effects for the non-quizzed categories. Hence, it appears that feedback about the performances in a category may have influenced predictions about future outcomes regarding that same category, but it did not cause any general shifts in strategy or attention (i.e., no generalization or transfer to non-quizzed categories). This pattern might provide some clues as to what people are really learning and applying when making predictions about quizzed categories. However, we will delay discussion of this issue until later in the paper.

Study 3

Studies 1 and 2 clearly illustrate that the mechanisms leading to SCEs are not easily mitigated. Repeated play and feedback across rounds did not appear to lead to a groundswell of eureka moments in which participants suddenly realized that they were not adequately factoring in their co-participant's level of knowledge. Let alone the fact that there was no groundswell of eureka moments, we were struck by the fact that, in Study 2, the shrinkage in the SCE for quizzed categories across rounds was paltry (see Table 3). For Study 3, we aimed to create paradigm conditions that were even more conducive for allowing people to learn to avoid SCEs. In keeping with our overall goals for this research—to see if and how people *learn their own way out* of egocentrism and SCEs—we continued to use natural feedback, rather than bogus feedback.

There were two basic changes in Study 3. First, in both Studies 1 and 2, the number of rounds or opportunities to receive information about a competitor may have been insufficient in providing participants the opportunity to learn their way out of egocentrism and SCEs. Thus, we extended the competition to six rounds instead of three. Second, perhaps the presence of two non-quizzed categories taxed participants' abilities to attend to, recall, and utilize appropriate information gained from experience and feedback from the quizzed categories. Thus, in Study 3, we completely removed the non-quizzed categories; instead, participants made predictions, answered questions, and received feedback about the same two categories across all six rounds. We believed these conditions were quite conducive for allowing participants to minimize egocentric weighting and SCEs across rounds.

[6] In hierarchical analyses similar to those described in Footnote 4, the self-knowledge estimates that were added in the second step again tended to account for significant variance ($p < .01$). The overall average R-squared change was 20% (and this value did not shift as a function of round or whether the category was quizzed).

Table 5
Mean Likelihood and Absolute Knowledge Judgments in Study 3 by Category Difficulty and Round ($N = 25$)

	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6
Likelihood Judgments						
Easy						
M	58.4*	59.4*	57.4*	60.0*	61.0*	62.3*
SD	14.3	14.9	15.0	17.2	16.5	20.4
Hard						
M	39.5*	41.6*	45.0*	47.1	49.2	52.8
SD	23.4	21.3	19.1	23.6	22.9	20.9
SCE						
M	18.9*	17.8*	12.4*	12.9*	11.8*	9.5*
SD	25.2	20.5	22.1	6.3	27.3	28.3
Absolute Knowledge Judgments						
Easy–Self						
M	4.28	4.79	4.96	4.88	5.00	5.20
SD	1.27	1.09	1.13	1.16	1.19	1.32
Easy–Other						
M	5.00	4.84	4.84	4.80	4.92	4.92
SD	0.91	0.80	0.89	1.04	0.81	1.15
Hard–Self						
M	2.60	2.56	2.60	2.60	2.60	2.68
SD	1.29	1.08	1.15	1.22	1.12	1.21
Hard–Other						
M	3.36	3.36	3.16	3.16	2.80	2.80
SD	1.11	1.41	1.18	1.43	1.26	1.19

Note. An asterisk in the *Hard* or *Easy* rows under *Likelihood Judgment* indicates that the mean was significantly different from 50% ($p < .05$). An asterisk in the *SCE* row indicates that the SCE was significantly different from 0 ($p < .05$). All absolute judgment values are significantly different from the midpoint (4) ($p < .05$), except for the mean for Easy–Self in Round 1, $t(24) = 1.1$, $p = .28$.

Method

The methods of Study 3 were identical to those of Study 2, with the following exceptions. Participants ($N = 25$) provided likelihood and knowledge estimates regarding the same two categories (instead of 4) per round. One category was easy (Pop Culture) and one was hard (South American Geography).[7] All participants also answered quiz questions and received full feedback for both of these categories in each of 6 (rather than 3) rounds. The design was a 2 (Category Difficulty) × 6 (Round) within subject factorial.

Results

Likelihood judgments

Mean likelihood judgments—as a function of category difficulty and round—are listed in Table 5. Preliminary analyses confirmed that, in Round 1, participants were significantly overoptimistic (compared to 50%) about a victory in easy categories, $t(24) = 2.3$, $p < .03$, and underoptimistic about a victory in hard categories, $t(24) = 2.9$, $p < .01$.

Next, we tested whether the magnitude of this SCE changed as a function of round. Given that there were 6 rounds, a linear trend analysis is more informative than a standard ANOVA. We examined linear trends for 3 separate sets of data across the 6 rounds: (1) difference scores—between the likelihood estimates for easy and hard categories—that reflect the magnitude of the SCE, (2) the likelihood estimates for the easy category, and (3) the likelihood estimates for the hard category. There was a trend for the SCE to shrink in size across rounds, although this effect did not reach significance, $\beta = -.13$, $F(1,148) = 2.4$, $p = .12$. This trend appeared to be entirely due to increases in likelihood estimates about the hard category across rounds, for which there was a significant linear trend, $\beta = .21$, $F(1,148) = 6.4$, $p < .02$. There was no significant linear trend for estimates about the easy category, $\beta = .08$, $F(1,148) = .97$, $p > .10$.[8] We also conducted analogous within subject tests of linear trends, where we ran the 3 regression analyses mentioned above separately for each participant and then averaged the resulting betas. The results were virtually identical to those reported above, where we found a significant trend in the hard category, but only directional trends for the SCE and easy category. In short, as can be seen in Table 5 and Fig. 2, people generally exhibited smaller SCEs across rounds (18.9% in Round 1 to 9.5% in Round 6), but this reduction in the bias seemed to be driven by increased optimism in the hard category.

Knowledge estimates

Table 5 also displays the means for knowledge judgments across rounds. Preliminary analyses on knowledge estimates revealed that, at Round 1, participants rated themselves as less knowledgeable than their competitor regarding the easy category, $t(24) = -2.5$, $p < .02$, and regarding the hard category, $t(24) = -2.7$, $p < .02$. When data for all rounds were collapsed, the former type of analysis (for the easy category) was non-significant, $t(24) = -.16$, $p > .1$, but the latter type

[7] Although it is always questionable to assume one stimulus can represent an entire conceptual category (Wells & Windschitl, 1999), we made this assumption here with the understanding that any results would need to be replicated with new categories (in Study 4).

[8] We conducted a follow-up test of whether the rate of the mitigation of bias was different for hard versus easy categories across rounds. First, to make the trends for the easy and hard categories line up directionally, we subtracted the hard-category estimates from 100. The easy and transformed hard estimates were submitted to a regression analysis that included category difficulty and round as predictors in the first step, then the interaction term in the second step. The interaction, which is the most important result here, was small in magnitude but significant, indicating that the rate at which bias was mitigated across rounds was greater for hard than for easy categories, $\beta = -.15$, $t(299) = -2.6$, $p < .02$ (R-squared change = 2%).

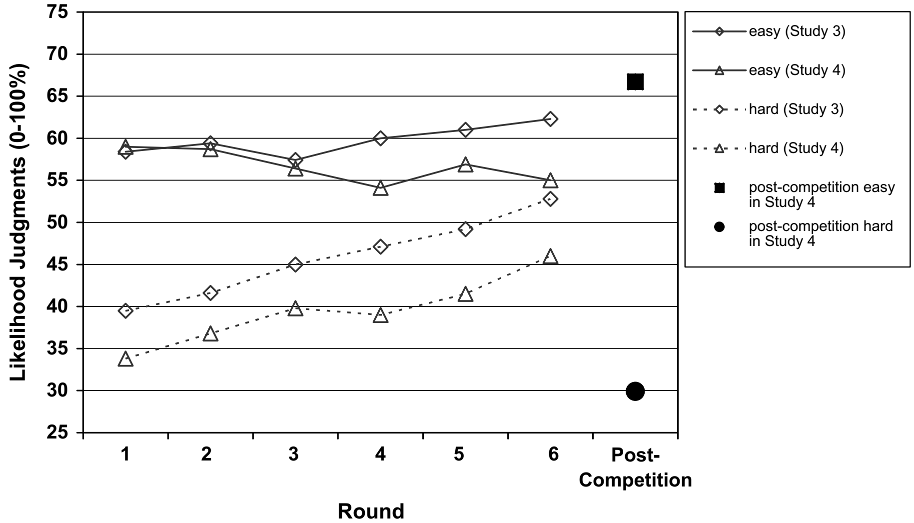

Fig. 2. Mean Likelihood Judgments in Studies 3 and 4.

of analysis (for the hard category) was significant, $t(24) = -2.1$, $p < .05$. These effects are somewhat subtle, however, and as a whole it is clear that participants generally believed that both the self and a competitor were very knowledgeable in the easy category and not very knowledgeable in the hard category.

The more critical analyses tested for linear trends (separately) across rounds for 4 sets of knowledge estimates: (1) self/easy, (2) competitor/easy, (3) self/hard, and (4) competitor/hard. Two of the four analyses revealed non-significant trends (ps $> .70$), except the first and fourth, which showed that participants' estimates of self-knowledge on the easy category increased across rounds, $\beta = .21$, $F(1, 148) = 6.9$, $p < .05$ and estimates of a competitor's knowledge on the hard category decreased across rounds, $\beta = -.73$, $F(1, 148) = 4.6$, $p < .05$.

Weighting of self-knowledge assessments versus competitor-knowledge assessments

As in Study 2, we conducted between-subjects regression analyses, treating probability estimates as the criterion variables and self- and competitor-knowledge ratings as the predictor variables. Self-knowledge estimates were better predictors of optimism in Round 1 (Mean $\beta = .81$) than were estimates of competitor knowledge (Mean $\beta = .03$). However, as is shown in Table 6, the disparity in beta weights between self- and competitor-knowledge decreased across rounds from a difference of .84 in Round 1 to .41 in Round 6. The tendency for this disparity to decrease in a linear fashion across Rounds 1–6 was significant based on a regression treating round as the unit of analysis, $\beta = -.89$, $t(5) = -3.8$, $p < .02$.[9]

Table 6
Results for path analyses from Study 3 ($N = 25$)

	β (Self)	β (Other)	Differential	r
Round 1	.81	.03	.84	.52
Round 2	.80	−.19	.61	.55
Round 3	.87	−.35	.52	.66
Round 4	.81	−.31	.50	.43
Round 5	.87	−.36	.51	.65
Round 6	.89	−.48	.41	.55

Note. The path analyses for Study 3 were conducted within category and across participants. The results were then averaged across categories. The differential column conceptually reflects the extent to which the self and other assessments have equivalent weight in predicting likelihood judgments (see note from Table 4 for additional information).

Discussion

It appears from Study 3 that at least under some conditions, SCEs may decrease in magnitude when people are given true feedback across multiple rounds of a competition. However, before discussing the underlying causes of this learning, we note that there are three reasons to replicate Study 3. First, the mitigation in the SCE was not statistically significant (perhaps because of the low sample size), and therefore a replication of the pattern would boost confidence that the mitigation effect was not due to random error. Second, in constructing the materials for Study 3, we selected only one easy and one hard category to represent larger classes of "easy" and "hard." We did this with the expectation that our results would need to be replicated with

[9] Our hierarchical regression analyses (see Footnotes 4 and 6) produced compatible findings. For analyses on data from Round 1, the R-squared change for entering self-knowledge estimates (after difference scores) was 40%, which reflects egocentric weighting. However, this value dropped across rounds, consistent with a reduction in egocentric weighting. For analyses on data from Round 6, the R-squared change for entering self-knowledge was only 11%.

different easy and hard categories (for a discussion of stimulus sampling, see Wells & Windschitl, 1999). Third, there was an asymmetry in how the SCEs were reduced, with participants increasing their optimism about hard categories over rounds but not changing their optimism about easy categories over rounds (and simply remaining overoptimistic even in Round 6) (see Footnote 8). This asymmetry could be important and is another reason for a replication. Study 4 was therefore designed as a replication of Study 3.

Study 4

In addition to replicating Study 3, we designed Study 4 to test the extent to which a decrease in egocentrism across the 6 rounds would reflect a level of insight that would be transferred and used when making subsequent judgments about different categories. More specifically, we tested whether, after 6 rounds of answering questions and receiving full feedback about performances in the same 2 trivia categories, participants would continue show reduced egocentrism and SCEs when judging the likelihood of outperforming their competitor on completely novel categories.

Method

Study 4 had 54 participants and was identical to Study 3 except for the following. First, the categories were Fast Food Chains (easy) and 50's Movies (hard). Second, when all 6 rounds were complete, participants also gave likelihood estimates about beating their competitor if they were to play 10 novel categories (5 easy, 5 hard; taken from Windschitl et al., 2003).

Results

Likelihood judgments

Mean likelihood judgments are listed in Table 7. As in Studies 1, 2, and 3, preliminary analyses confirmed that, in Round 1, participants were significantly overoptimistic (compared to 50%) about a victory in easy categories, $t(53) = 3.9$, $p < .01$, and underoptimistic about a victory in hard categories, $t(53) = -5.8$, $p < .01$. Next, an examination of linear slopes (same as for Study 3) revealed significant trends for the SCE to shrink in size across rounds, $\beta = -.19$, $F(1, 327) = 12.6$, $p < .01$, from 25.5% in Round 1 to 9.2% in Round 6. As in Study 3, a key part of this trend was a significant increase in optimism about the hard category across rounds, $\beta = .17$, $F(1, 327) = 9.8$, $p < .01$. There was a non-significant trend for easy category judgments to become more *pessimistic*, $\beta = -.08$, $F(1, 327) = 2.3$, $p = .13$.[10] As in Study 3, we also conducted analogous within subject tests of the 3 linear trends described above. The respective results were comparable to those reported above and, in fact, even stronger than the between subject effects because all 3 trends were significant. See Fig. 2 for a visual depiction of these patterns.

Knowledge estimates

Table 7 also displays the means for knowledge judgments across rounds. Participants rated themselves as roughly equally knowledgeable as their competitor in the easy category at Round 1, $t(53) = -1.4$, $p > .10$, and with all rounds collapsed, $t(53) = -1.9$, $p > .05$. Participants rated themselves as less knowledgeable than their competitor in the hard category at Round 1,

Table 7
Mean Likelihood and Absolute Knowledge Judgments in Study 4 by Category Difficulty and Round ($N = 54$)

	Round						"Post-Comp
	1	2	3	4	5	6	
Likelihood Judgments							
Easy							
M	59.0*	58.7*	56.4	54.1	56.9	55.0	66.7*
SD	20.7	15.9	15.2	16.9	16.7	14.7	16.1
Hard							
M	33.8*	36.8*	39.8*	39.0*	41.5*	46.0	29.9*
SD	20.8	21.9	21.8	21.1	19.6	16.7	14.3
SCE							
M	25.5*	21.8*	16.8*	15.2*	15.4*	9.2*	36.8*
SD	23.3	21.3	25.6	27.2	26.1	27.9	23.1
Absolute Knowledge Judgments							
Easy–Self							
M	4.89	4.96	4.98	5.07	5.23	5.00	—
SD	0.93	1.04	1.13	1.10	1.07	1.04	—
Easy–Other							
M	5.07	5.02	5.24	5.16	5.24	5.22	—
SD	0.99	1.14	1.09	1.10	1.03	1.08	—
Hard–Self							
M	2.25	2.00	1.95	1.98	1.95	2.05	—
SD	1.16	1.26	1.03	1.06	1.09	1.15	—
Hard–Other							
M	2.76	2.45	2.56	2.41	2.45	2.25	—
SD	1.08	1.25	1.27	1.18	1.15	1.16	—

Note. An asterisk in the *Hard* or *Easy* rows under *Likelihood Judgment* indicates that the mean was significantly different from 50% ($p < .05$). An asterisk in the *SCE* row indicates that the SCE was significantly different from 0 ($p < .05$). All absolute judgment values are significantly different from the midpoint (4) ($p < .05$). In the *Post-Comp* column are the mean likelihood estimates (and SCE) for easy/hard categories that were calculated from participants' post-competition estimates that were taken after all 6 rounds of competition. Asterisks in this column indicate these are significantly different from 50% (and 0) ($p < .05$).

[10] Again, because of the potential category difficulty asymmetry, we used a regression approach to test for a Round × Category Difficulty interaction (see Footnote 8). This analysis did not reveal a significant interaction, $\beta = .05$, $t(649) = 1.4$, $p > .15$ (R-square = .3%).

Table 8
Results for path analyses from Study 4 ($N = 54$)

	β (Self)	β (Other)	Differential	r
Round 1	.74	−.08	.66	.79
Round 2	.98	−.39	.59	.82
Round 3	1.10	−.62	.48	.78
Round 4	1.22	−.79	.43	.85
Round 5	1.01	−.52	.49	.83
Round 6	.76	−.43	.33	.84

Note. The path analyses for Study 4 were conducted within category and across participants. The results were then averaged across categories. The differential column conceptually reflects the extent to which the self and other assessments have equivalent weight in predicting likelihood judgments (see note from Table 4 for additional information).

$t(53) = -3.4$, $p < .01$, and with all rounds collapsed, $t(53) = -3.6$, $p < .01$. All four linear trends (same as those in Study 3) were not significant (all $ps > .09$).

Weighting of self-knowledge assessments versus competitor-knowledge assessments

The results of regression analyses treating probability estimates as the criterion variable and self-knowledge and competitor-knowledge ratings as the predictor variables are shown in Table 8. As was the case for Study 3, there was a large disparity in beta weights at Round 1 (favoring the self-knowledge estimates) and this disparity decreased across rounds. The tendency for this disparity to decrease in a linear fashion across Rounds 1–6 was significant based on a regression treating round as the unit of analysis, $\beta = -.92$, $t(5) = -4.6$, $p < .01$.[11]

Generalization to novel categories

The results for Study 4 reported thus far suggest that people became less egocentric across rounds. Was this reduction in egocentrism and SCEs due to a form of insight that participants would transfer to their likelihood judgments about new categories? Or, was the learning category specific? At the end of the study and after having received full feedback about Round 6 performances, participants made likelihood estimates about winning in 10 novel categories (5 easy, 5 hard). Participants were much more optimistic about winning easy categories ($M = 66.7$, $SD = 14.3$) than hard ones ($M = 29.9$, $SD = 14.3$), $t(53) = 11.9$, $p < 01$. The SCE—which shrunk across the six rounds—experienced a rebound when novel categories were introduced. The magnitude of the SCE on these novel categories ($M = 36.8$) is approximately the same as the magnitude observed in Round 1 of Study 1 ($M = 32.3$) and in Study 3 of Windschitl et al. (2003) ($M = 44.0$). Therefore, it appears that participants in Study 4 were not applying a special insight that helped them avoid SCEs when making likelihood judgments about the new categories. The utilization of full feedback in Rounds 1–6 was category specific rather than transferable. This finding is consistent with other research showing that feedback/experience can mitigate various judgment biases (e.g., the endowment effect, winner's curse) in specific circumstances, but does not produce generalized or transferable effects (Kagel & Levin, 1986; Van Boven et al., 2003).

Discussion

Although we inserted a new easy and hard category for Study 4 (relative to Study 3), the results of Study 4 replicated the pattern in Study 3: egocentrism and the SCE decreased over the course of 6 rounds of play and feedback. However, an additional set of likelihood judgments from Study 4 showed that the reduction in the SCE was limited to the two categories about which participants were quizzed and received feedback.

General discussion

Does repeated experience with the same competitive context, the same judgment task, and the same competitor push participants into learning their way out of bias? As a whole, the results from four studies lend support to the idea that repeated experience can lead to a reduction in egocentrism and SCEs. However, the results also suggest that—at least within the trivia paradigm—SCEs are not quickly reduced and that the eventual reduction was not due to insights that were readily generalized or transferred.

In Study 1, participants answered questions, made optimism judgments, and received feedback in 3 rounds of competition. Each round involved a new set of categories. Egocentrism and the level of SCEs remained high across each round in Study 1. In Study 2, we enhanced the feasibility of learning in our participants by reducing the overall number of categories involved in the competition (reducing cognitive overload), increasing the number of questions per category (increasing the quantity/quality of feedback), and having participants answer questions in the same two categories across all three rounds (increasing the compatibility of past and future rounds). These methodological changes produced only modest evidence of reduced egocentrism and SCEs. In Study 3, participants made optimism judgments, answered questions, and received feedback about the same two trivia categories across 6 rounds of competition. The results revealed trends in which egocentric weighting and SCEs shrunk from Rounds 1 to 6. In

[11] Again, our hierarchical regression analyses (see Footnotes 4, 6, and 9) produced compatible findings. For analyses on data from Round 1, the R-squared change for entering self-knowledge estimates (after difference scores) was 22%. For analyses on data from Round 6, it was only 8%.

Study 4, we replicated the results of Study 3 using two new categories. In both studies, the reduction in SCEs was primarily due to a reduction in underoptimism for hard categories, rather than a reduction in overoptimism for easy categories (this issue is discussed later). Another important finding from Study 4 was that the reduction of SCEs that occurred from Round 1 to 6 was restricted to the categories in which participants were repeatedly quizzed and received feedback. Any insights that led to reductions in participants' bias were not transferred to their predictions about novel categories.

Earlier, we mentioned that one other study has examined SCEs in repeated-play contexts (Study 1 of Moore & Cain, 2007). Several critical differences exist between the methods of that study and our studies, but it is nevertheless noteworthy that the findings from their study are most compatible with the findings from our Study 1. In their study, each relevant round involved one quiz, which was either easy or hard, and the critical dependent variable (with respect to the current issues) was participants' estimates of their percentile rank on quiz performance among the competitors. Although participants received full feedback after each of numerous rounds, the extent to which they exhibited SCEs at the end of the experimental session was the same as it was at the beginning. Like our Study 1, in which categories were not repeated across rounds, the specific quizzes in the Moore and Cain (2007) study were never repeated. We suspect this feature of the Moore and Cain study is a key reason why those participants showed no reduction in SCEs across rounds.

What were participants learning?

In general, our studies provide evidence that participants used feedback and experience in some manner to reduce biases in optimism. But what did participants actually learn in our studies? Further, was this learning superficial, or deep and transferable? We discuss several possibilities.

One possibility is that the salience of a competitor may have increased with experience/feedback. Previous research has indicated that the salience of, or direct contact with, a comparison referent can influence comparative bias (Alicke et al., 1995), and research on focalism indicates that referents (rather than targets) are routinely underweighted when comparative or likelihood judgments are made (Chambers & Windschitl, 2004; Eiser et al., 2001; Giladi & Klar, 2002; Kruger et al., in press; Moore & Kim, 2003; Pahl & Eiser, 2007; Windschitl et al., 2003). Hence, repeated experience and feedback about a competitor might have influenced our participants' tendencies to give inadequate attention to their competitor.

A second possibility is that participants may have learned to regress their likelihood estimates toward chance levels (50%). This decision to regress estimates toward 50% may have been due to some realization that the competition itself includes chance fluctuations in performances, or the decision to regress may have been a reaction toward expectancy violations in performances (i.e., losing in an easy category, winning in a hard category). In essence, participants may have learned to be strategically more conservative in their judgments.

Although these two accounts might help explain some of the reduction in SCEs, we believe both accounts have limited applicability. Recall that at the end of Study 4, when participants were asked to judge the likelihood of winning 10 new categories, they were again extremely overoptimistic about winning easy categories and extremely underoptimistic about hard ones (see Table 7 and Fig. 2). If increased salience of the competitor (or his/her knowledge and performances) was a major reason why SCEs shrunk across rounds, we would expect this increased salience to have had a more substantial impact that it did on the judgments about the novel categories. Similarly, if there was a strategic trend toward greater conservatism, we would expect this strategy to also be used in novel category judgments.

Furthermore, the fact that SCEs enjoyed a robust rebound at the end of Study 4 does not bode well for the possibility that the reduction in SCEs in any of the studies was due to a deeper realization about past egocentric tendencies (i.e., "When I gauge my likelihood of winning, I need to start thinking about not just my own level of knowledge but also my competitor's level of knowledge"). Across the four studies, it appears that participants were not gaining deep-level insight into egocentric tendencies, or at least that the insights were not transferable unless the surface similarities across rounds were pristine (Bassok, 2003).

Although we do not wish to completely rule out increased salience, increased conservatism, and deep insight as at least minor contributors to the reduction of SCEs in our studies, we believe that the bulk of the reduction is due to the direct informational content of the feedback. As emphasized above, the reduction in SCEs occurred when participants were making likelihood judgments regarding the same categories about which they had received feedback.

With this observation in mind, one might assume that feedback changed people's estimates of what their competitor knew about a critical category, which thereby influenced likelihood judgments. For example, participants may have learned that their competitor was not as knowledgeable about hard categories as they had initially believed. Data from knowledge estimates in Study 3 illustrates this possibility (see Table 5). At Round 1 of that study, participants indicated that their competitors were more knowledgeable than the self regarding a hard category. This difference disappeared by Round 6, as did people's underoptimism about winning the hard categories. Although this might appear to support the idea that changes in people's estimates of their competitor's knowledge was crucial, these changes cannot explain much of the general reduction in SCEs. In Study 4, for example, participants' beliefs about their competitor's knowledge (or their own knowledge) on the hard category did not shift significantly, even though there was a substantial reduction in underoptimism for this category. Also, for every study, participants were overoptimistic about winning easy categories even though they estimated that they had less or the same level of knowledge as their competitor. This suggests that differences in how much knowledge people believed the self and the competitor had could only provide a *partial* account for SCEs and their reductions.

We believe the account that is most plausible as the major contributor to the reduction in SCEs is the differential-confidence account (Chambers & Windschitl, 2004; Kruger et al., in press). This account suggests that in Round 1 of each study, although participants knew that they and their competitor had little knowledge about hard categories and much knowledge of easy ones, they were more confident about their self-assessments than their competitor-assessments. Hence, the self-assessments were given more weight when likelihood judgments were made. However, as participants became more confident in their assessments of the competitor's knowledge for specific categories (i.e., the repeated categories in Studies 2–4), the differential weighting became less extreme. This rather rational account of SCEs and their reduction can also account for the rebound in SCEs in Study 4. Namely, participants had never received feedback about the competitor's knowledge of the new categories, and were therefore again more confident about their self-assessments than about their competitor-assessments. In sum, we suspect that there are multiple ways that feedback from repeated plays can mitigate SCEs. However, the pattern of results for the present studies suggests that—in this paradigm—the processes described by the differential-confidence account were most critical.

With that said, we should note that had we asked for a different form of absolute judgment—such as point predictions of how many items the self and the competitor would answer correctly—a differential-regression account might have received more support. A differential-regression account suggests that because participants have less knowledge of the competitor than of the self, their absolute judgments about the competitor would tend to be less extreme (i.e., more regressive) than their absolute judgments of the self (see Chambers & Windschitl, 2004; Kruger et al., in press; Moore, 2007; Moore & Cain, 2007; Moore & Small, 2007). Had we asked for point predictions, rather than knowledge estimates, it might have been the case that, as more feedback was received across rounds, the knowledge gap and regression gap would have substantially narrowed—coinciding and possibly explaining the reduction in bias in likelihood judgments. Indeed, recent research suggests that judgments made on common-rule scales (such as a point prediction) are often more sensitive to such differential-regression effects than are judgments with more subjective interpretations (such as a 1–7 knowledge scale; see Moore, 2007; see also Burson & Klayman, 2006). For the present research, we chose to solicit knowledge estimates rather than point predictions because they seemed most psychologically relevant to the likelihood question faced by our participants or by others assessing the likelihood of winning a competition. For example, when a student thinks about whether he/she will outperform a classmate on a chemistry test, the questions of how knowledgeable the self and a friend are on the subject matter would presumably be more salient and relevant to the student than the questions of how many points the self and the friend will score correctly.[12] In short, the issue of whether differential-regression accounts would be better supported when point predictions are used is open for future research to address, but we do not think this possibility necessarily impinges on our conclusion that the present findings are consistent with a differential-confidence interpretation.

Asymmetrical influence of category difficulty

An additional issue to address is why there appeared to be an asymmetrical influence on likelihood estimates based on category difficulty in Studies 2–4. The appearance of bias in these estimates seemed to decrease more across rounds in hard categories than in easy categories—although this differential in the decrease was significant only in Study 3 (see Footnotes 8 and 10 for data regarding the category difficulty $\times$ round interactions in Studies 3 and 4). It is possible that the asymmetry was partially due to motivated reasoning or a desire to view the self in a positive light (for reviews, see Krizan & Windschitl, 2007b; Kunda, 1990). For instance, participants may have been more receptive and attentive to feedback that was better than expected (as would generally be the case for hard categories, where initial expectations

[12] It seems that point predictions might be more psychologically relevant when participants have already performed a task and may enumerate items on which they were successful. In predictions about the future, however, such enumeration seems less relevant. Also, we think it is noteworthy that in recent research using a trivia paradigm in which participants were asked to make point predictions and likelihood judgments *after* taking a quiz, we still found robust evidence for differential weighting—above and beyond any influence of differential regression (Windschitl, Rose, Stalkfleet, & Smith, 2007).

were low) as opposed to worse than expected (as would be the case for easy categories) (see Dawson, Gilovich, & Regan, 2002; Ditto, Scepansky, Munro, Apanovitch, & Lockhart, 1998). Participants may have also made differential attributions for positive feedback (e.g., "I am actually pretty good") versus negative feedback (e.g., "My competitor got lucky"). These attribution differences may have generally influenced the direction of future optimism in an upward direction, which would partially mitigate any tendency to be less overoptimistic about easy categories.

Although each of the above suppositions are plausible, we also suggest the asymmetry may be an artifact of participant expectations and the differences in magnitude/extremity between likelihood judgments for easy versus hard categories. For instance, participants' likelihood estimates about a victory in hard categories were much more extreme and subsequently more susceptible to experiencing violations in expectations after gaining experience and feedback. Across all studies, the average likelihood estimate for a victory in Round 1 for easy categories was 60.0 %, whereas the average estimate for hard categories was 32.5%. If we assume that 50% is the objectively unbiased point, this suggests that participants were more certain about losing in hard categories ($50 - 32.5 = 17.5$ points away from 50%) than they were about winning in easy categories ($60.0 - 50 = 10$ points away from 50%). These initial expectations (in Round 1) may have influenced the asymmetry in judgments based upon category difficulty, simply because expectations were violated to a greater degree for hard categories (because they were more extreme and further away from a 50/50 chance estimate) and because participants had more "room" to adjust estimates toward the unbiased point (i.e., hard category estimates were further away from 50% initially). Currently, then, there are both motivational and non-motivational explanations for the asymmetry in how optimism changed for easy and hard categories. Only future research could disentangle those explanations and yield firmer conclusions about the issue.

Conclusions

This work has demonstrated that people can use natural feedback in a way that reduces overoptimism and underoptimism (i.e., SCEs) in repeated competitions. However, the work is perhaps more profound in illustrating that these reductions in bias will come slowly and only under limited circumstances. We suggest that this is due to two reasons. The first reason is that the overoptimism and underoptimism embodied in the SCE studies in this paradigm were based in part on a generally rational process of weighting knowledge assessments according to the confidence with which they are held. A second reason is that any insights or changes in strategies that people may have developed regarding one category were not readily transferred to judgments about other categories. Hence, determining whether repeated plays will result in better calibrated optimism within other competitions (e.g., employment, athletics) will likely depend on (1) whether people have much more confidence on how a shared circumstance will affect the self rather than a competitor and (2) whether the repeated tasks are identical or, at least, obviously similar in nature.

Appendix 1

List of categories for Study 1

Block 1
1. Russian literature (H)*
2. South American geography (H)
3. Adam Sandler movies (E)
4. Famous composers (H)
5. Home insurance facts (H)*
6. Pop music (E)*
7. Brands of alcohol (E)
8. TV sitcoms (E)*
9. Eastern philosophy (H)
10. Famous cartoon characters (E)

Block 2
1. State Capitols (E)
2. 50's Movies (H)*
3. History of Mesopotamia (H)
4. 19th Century French painting (H)
5. US Geography (E)*
6. European politics (H)
7. Clothing fads (E)
8. Driving rules (E)*
9. Famous rivers (H)*
10. Current events (E)

Block 3
1. Fast food chains (E)*
2. Celebrities (E)
3. Latin American history (H)*
4. Baroque music (H)
5. Using a personal computer (E)
6. Ancient civilizations (H)
7. Rock 'n' Roll (E)
8. Dates in history (H)*
9. Vegetation of the Amazon (H)
10. Pop culture (E)*

Note. Level of category difficulty is indicated in parentheses: E = easy; H = hard. Categories selected as the subset from which questions were asked are indicated by an asterisk (*).

Appendix 2

Sample questions for Study 1

Hard category multiple-choice question
Construction of the Panama Canal began in what year?
(a) 1963
(b) 1904
(c) 1946
(d) 1852

Easy category multiple-choice question
Long John Silver's specializes in what type of fast food?
(a) Seafood
(b) American food
(c) Mexican
(d) Italian

Numeric tie-breaker question
Approximately how many people watched the finale of Mash? _________

References

Alicke, M., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vrendenburg, D. S. (1995). Personal contact, individuation, and the better-than average effect. *Journal of Personality and Social Psychology, 5*, 804–825.

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes, 39*, 133–144.

Ball, S. B., Bazerman, M. H., & Carroll, J. S. (1991). An evaluation of learning in the bilateral winner's curse. *Organizational Behavior and Human Decision Processes, 48*, 1–22.

Bassok, M. (2003). Analogical transfer in problem solving. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 343–369). New York, NY, US: Cambridge University Press.

Bazerman, M. H., & Samuelson, W. F. (1983). I won the auction but don't want the prize. *Journal of Conflict Resolution, 27*, 618–634.

Buckingham, J. T., & Alicke, M. D. (2002). The influence of individual versus aggregate social comparison and the presence of others on self-evaluations. *Journal of Personality and Social Psychology, 83*, 1117–1130.

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

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.

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

Chambers, J., & 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., 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.

Cox, J. C., & Grether, D. M. (1996). The preference reversal phenomenon: Response mode, markets, and incentives. *Economic Theory, 7*, 381–405.

Dawson, E., Gilovich, T., & Regan, D. T. (2002). Motivated reasoning and performance on the Wason Selection Task. *Personality and Social Psychology Bulletin, 28*, 1379–1387.

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology, 75*, 53–69.

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

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*, 5–17.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85*, 395–416.

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.

Fischer, G. W. (1982). Scoring-rule feedback and the overconfidence syndrome in subjective probability forecasting. *Organizational Behavior and Human Performance, 29*, 352–369.

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387–396.

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.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*, 332–340.

Kagel, J. H., & Levin, D. (1986). The winner's curse and public information in common value auctions. *American Economic Review, 76*, 894–920.

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.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Nizam, A. (1998). *Applied regression analysis and other multivariable methods* (3rd ed.). Pacific Grove, CA: Duxbury Press.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*, 147–162.

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

Krizan, Z., & Windschitl, P. D. (2007b). The influence of outcome desirability on optimism. *Psychological Bulletin, 133*, 95–121.

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., Windschitl, P. D., Burrus, J., Fessel, F., & Chambers, J. R. (in press). The rational side of egocentrism in social comparisons. *Journal of Experimental Social Psychology*.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.

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

McClosky, M. (1983). Naïve theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 254–324). Hillsdale, NJ: Lawrence Erlbaum.

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 & 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 judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology, 92*, 972–989.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 510–520.

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

Radhakrishan, P., Arrow, H., & Sniezek, J. A. (1996). Hoping, performing, learning and predicting: Changes in the accuracy of self-evaluations of performance. *Human Performance, 9*, 23–49.

Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 436–450.

Reed, S. K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 106–125.

Renner, C. H., & Renner, M. J. (2001). But I thought I knew that: Using confidence estimation as a debiasing technique to improve classroom performance. *Applied Cognitive Psychology, 15*, 23–32.

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

Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Test difficulty and judgment bias. *Memory and Cognition, 22*, 63–69.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024–1037.

Van Boven, L., Loewenstein, G., & Dunning, D. (2003). Mispredicting the endowment effect: Underestimation of owners' selling prices by buyer's agents. *Journal of Economic Behavior & Organization, 51*, 351–365.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115–1125.

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*, 349–408.

Windschitl, P. D., Rose, J. P., Stalkfleet, M., & Smith, A. S. (2007). Are people excessive or judicious in their egocentrism? A modeling approach to understanding bias and accuracy in people's optimism within competitive contexts. Submitted for publication.

Zakay, D. (1992). The influence of computerized feedback on overconfidence in knowledge. *Behavior & Information Technology, 11*, 329–333.

Zechmeister, E. B., Rusch, K. M., & Markell, K. A. (1986). Training college students to assess accurately what they know and don't know. *Human Learning: Journal of Practical Research and Applications, 5*, 3–19.