University of St. Thomas, Minnesota
UST Research Online

Marketing Faculty Publications   Marketing

2006

Accuracy, Error, and Bias in Predictions for Real Versus Hypothetical Events

David A. Armor
Yale University

Aaron M. Sackett
University of St. Thomas, Minnesota, sackett@stthomas.edu

Follow this and additional works at: http://ir.stthomas.edu/ocbmktgpub

Part of the Business Administration, Management, and Operations Commons, and the Marketing Commons

Accuracy, Error, and Bias in Predictions for Real Versus Hypothetical Events

David A. Armor and Aaron M. Sackett
Yale University

Participants made predictions about performance on tasks that they did or did not expect to complete. In three experiments, participants in task-unexpected conditions were unrealistically optimistic: They overestimated how well they would perform, often by a large margin, and their predictions were not correlated with their performance. By contrast, participants assigned to task-expected conditions made predictions that were not only less optimistic but strikingly accurate. Consistent with predictions from construal level theory, data from a fourth experiment suggest that it is the uncertainty associated with hypothetical tasks, and not a lack of cognitive processing, that frees people to make optimistic prediction errors. Unrealistic optimism, when it occurs, may be truly unrealistic; however, it may be less ubiquitous than has been previously suggested.

Keywords: optimism, optimistic bias, prediction, accuracy, construal level theory

How accurate are people's visions of the future? The general consensus, which may be gleaned from even a brief perusal of research on personal forecasts, is that people are not very accurate at all. To date, several hundred studies have shown that people's predictions tend to be excessively and unrealistically optimistic (Weinstein, 1998; for reviews, see Armor & Taylor, 1998, 2002; Buehler, Griffin, & Ross, 2002; Helweg-Larsen & Shepperd, 2001). To be sure, some degree of error in personal prediction is to be expected. As Yogi Berra famously, if mythically, aphorized, "It's tough to make predictions, especially about the future."[1] However, people's predictions appear to be prone not only to *error*—random deviations and distortions that would be expected to balance out over the long run—but to consistent and pervasive *bias*. The common conclusion from studies of personal forecasts is that people expect that their futures will be more pleasant and less painful than they have any right to expect them to be.

Examples of optimistic biases in personal predictions are not difficult to find, whether in real life or in the relatively artificial confines of the research laboratory. Students, for example, tend to overestimate how well they will do on exams (Shepperd, Ouellette, & Fernandez, 1996) and to underestimate how long it will take for them to complete their assignments (Buehler, Griffin, & Ross, 1994). Undergraduate and professional (Master of Business Administration) students tend to overestimate their prospects for success on the job market (Hoch, 1985; Shepperd et al., 1996), and even gainfully employed professional financial analysts tend to overestimate corporate earnings (Calderon, 1993; Lim, 2001). A great deal of research has also shown that people of all ages and backgrounds tend to overestimate how likely they are to experience a wide variety of positive outcomes, and to underestimate how likely they are to experience an even wider variety of negative outcomes relative to other people (e.g., Weinstein, 1980, 1987; see also Perloff & Fetzer, 1986; Quadrel, Fischhoff, & Davis, 1993; cf. Heine & Lehman, 1995; for a review, see Helweg-Larsen & Shepperd, 2001).

Despite the prevalence of these demonstrations, which give the impression that optimistic biases may be nearly unavoidable (see, e.g., Weinstein & Klein, 1995), there is a growing body of research suggesting that this portrait of unwavering optimism may be overly simplified (see Armor & Taylor, 1998, 2002; Shepperd, Sweeny, & Carroll, 2006). There are at least two reasons for this. First, research that provides a clear criterion for evaluating prediction accuracy (e.g., by comparing predictions to attained outcomes) has shown that people's predictions often may be optimis-

[1] Although commonly attributed to Berra, this statement may be more a part of the Hall of Fame catcher's legend than of his loquacious history; Berra himself admitted "I really didn't say everything I said" (Berra, 1998). Near paraphrases have also been attributed, with comparable frequency, to Nobel Laureate Niels Bohr.

tically biased without being entirely unrealistic. Research on the planning fallacy, for example, has shown that whereas people's task completion estimates tend to be overly optimistic (in that people expect to be able to do more in less time than they are typically able to do), these same estimates tend to be quite highly correlated with actual task completion times (correlations between predicted and actual completion times in the initial studies of the planning fallacy ranged from .36 to .81; see Buehler et al., 1994). Thus, although biased, these optimistic predictions do not appear to be completely divorced from reality. One of the goals of the present research, then, is to provide a clear criterion against which the accuracy of individual predictions can be assessed.

Second, people do not appear to be indiscriminately optimistic but to be sensitive to the context in which they make their predictions (Armor & Taylor, 1998, 2002; Shepperd et al., 2006). There is an accumulating body of research suggesting that people may be more likely to be overly optimistic in situations in which it is less likely that the accuracy of their predictions will be tested or challenged or in situations in which the consequences of being inaccurate are expected to be less severe. For example, whereas students tend to overestimate how well they will do on their exams, this optimism is most prominent (and most unrealistic) when those exams are some time away; optimistic biases appear to be much less pronounced as the "moment of truth" draws near (Gilovich, Kerr, & Medvec, 1993; Shepperd et al., 1996, Study 2). Similarly, whereas several studies have found that students overestimate how much they will earn in their first job, students tend to see their job prospects in more modest terms—and, indeed, accurately estimate their starting salary—as they approach the market for real (Shepperd et al., 1996, Study 1). Even Calderon's (1993) analysis of professional analysts' financial predictions revealed a systematic decline in optimistic bias as a function of the proximity of the forecast date to the realization date (although optimistic bias was evident even among the most proximal predictions).

On the basis of results such as these, Armor and Taylor (1998, 2002) have argued that people are not indiscriminately optimistic, but are "situated optimists" and that the expression of optimistic biases is not invariant within persons but rather is largely dependent on the social and psychological context in which people find themselves.[2] The notion that optimistic biases are not inherent within persons but situated within contexts raises a number of important theoretical questions—namely, when and why do people tend to be unrealistically optimistic?—but also raises the question of whether the prevalence of unrealistic optimism may be over- or underestimated as a function of the experimental setting in which unrealistic optimism is typically studied (for similar arguments, see Gilovich et al., 1993, p. 559). In other words, it is not clear whether the seemingly common expressions of unrealistic optimism reflect a general tendency for people to be optimistic or an equally general tendency for psychologists to measure optimism in settings in which unrealistic optimism may be especially likely to be expressed.

The Social Psychology of the Unrealistic Optimism Experiment

In the early 1960s, Orne (1962) published an influential critique suggesting that psychologists—social psychologists in particular—should pay careful attention to the psychology experiment as

a "social situation" and that psychologists should be mindful of the effects that these situations have, above and beyond any effects of the experimental manipulations, on the thoughts, feelings, and observable behaviors of research participants. Tetlock's critique of decontextualized research on judgmental errors and biases makes similar arguments (e.g., Tetlock, 1992; Tetlock & Lerner, 1999; see also Schwarz, 1994).

Although a tremendous number of studies have shown that people tend to be excessively optimistic, a close look at the contexts in which these data have been collected reveal that the overwhelming majority of these studies have been conducted in a manner that may be especially conducive to the expression of unrealistic optimism. In most studies, accuracy incentives are either minimal or absent altogether, as participants are often asked to make predictions anonymously and without concern about possible consequences of inaccurate predictions. In many cases the outcome of participants' predictions simply cannot be known—by either the researcher or the person making the prediction—and, as a consequence, prediction accuracy itself is often impossible to verify. In the modal study, for example, young, healthy college students might be asked of their chances, relative to the chances of other young, healthy college students, of one day succumbing to cancer, of needing dentures, or of being fired from a job for which they have yet to be hired. Although these may all be meaningful questions, and although arguments can be made about their relevance to respondents' current behavior (e.g., decisions to smoke, eat sweets, or to develop the "seven coveted habits of highly effective employees"), these outcomes are also likely to be quite far removed from most participants' current concerns. To the extent that research participants are asked to estimate the probability of hypothetical outcomes or outcomes that are unimaginably distant, it seems possible that they may be aware of the fundamentally unverifiable nature of their predictions and that they may, knowingly or unknowingly, alter their predictions accordingly.

Of course, not all studies of unrealistic optimism have involved predictions for entirely hypothetical outcomes. Research on the planning fallacy, for example, has asked participants to make predictions about tasks they knew they would be completing (see Buehler et al., 2002, for review), and at least one study (Camerer & Lovallo, 1999) has found evidence of unrealistic optimism even when participants could (and did) lose money as a function of their overly optimistic self-assessments. It is not entirely clear, however, to what extent participants in these studies may have been even more optimistic—and potentially more *unrealistically* optimistic—had predictions been made without the expectation that predictions would be verified.

---

[2] The notion that optimism is situated does not deny that there are meaningful individual differences in people's tendency to be optimistic (see, e.g., Norem & Cantor, 1986; Scheier, Carver, & Bridges, 1994). However, studies examining the relationship between optimism as a general trait and specific instantiations of optimism in individual predictions reveal little correspondence between the two (Armor & Taylor, 1998; Buehler & Griffin, 2003). Our focus, therefore, is on outcome-specific expectations and not more generalized expectancies captured by measures of dispositional optimism.

Predictions for Real Versus Hypothetical Events

The present studies examine the accuracy of predictions people make for tasks that they do or do not expect to complete. Several lines of research seemed relevant, a priori, as sources of hypotheses, including Gilovich et al.'s (1993) research on the effects of temporal distance on subjective confidence; Trope and Liberman's (e.g., 2003; Liberman & Trope, 1998) research on construal level theory; Sedikides, Herbst, Hardin, and Dardis's (2002) work on accountability and self-evaluation; and Shepperd et al.'s (1996) research on the effects of expecting proximate feedback on pessimistic "bracing for the worst."

In Gilovich et al.'s (1993) experiments, people were found to predict, optimistically, that they would perform well above average on tasks that were described as being temporally distant (i.e., when the opportunity to complete these tasks was weeks or months away), but the extent to which people showed this effect of predicting "better than average" performance was markedly reduced when the tasks were described as being temporally proximate (i.e., when the opportunity to complete these tasks was only moments away). Although Gilovich et al. (1993) emphasized that their hypotheses and data pertained to temporal distance, and not to the real versus hypothetical status of the event in question (see, e.g., Gilovich et al., 1993, p. 552), there is reason to suspect that the two variables may have similar effects on people's predictions. According to Trope and Liberman's (2003) construal level theory, a variety of manipulations—including temporal distance and the real versus hypothetical status of events—may be functionally interchangeable as manipulations of "psychological distance." Insofar as hypothetical tasks may be seen as being infinitely temporally distant, we hypothesize that people will have considerably greater leeway to make overly optimistic predictions when the task in question is hypothetical as opposed to real.

Two other lines of work suggest that, regardless of psychological distance, the possibility of receiving evaluative feedback may be the critical independent variable (Sedikides et al., 2002; Shepperd et al., 1996; see also Shepperd, Grace, Cole, & Klein, 2005; Taylor & Shepperd, 1998). The potential for feedback—whether real or imagined, public or private, externally provided or internally derived—is one of the characteristics that differentiate real tasks from hypothetical ones. When tasks are real, one can (and often will) learn how well one has performed on that task, but when a task is hypothetical, opportunities for these kinds of reality checks are largely absent. However, studies demonstrating the effects of feedback expectations, which generally show a reduction of optimism, have focused largely on assessments made after participants had completed a critical task (e.g., estimates of how well one has performed on a task that one has just completed; Sedikides et al., 2002; Shepperd et al., 1996, 2005; see also Sniezek, Paese, & Switzer, 1990). Studies that have examined the effect of manipulated feedback expectations on pretask predictions have yielded inconsistent results, sometimes showing no effect (e.g., Study 2 of Buehler et al., 1994) and sometimes showing a reduction in optimism (e.g., Taylor & Shepperd, 1998). It is thus not clear that the expectation of feedback is a necessary precondition for a reduction in pretask optimism.

In the studies reported here, we examine the degree of accuracy, error, and bias in predictions people make about their performance on tasks that they believe are either real (i.e., tasks that they will

soon be completing) or hypothetical (i.e., tasks that they do not think they will complete). We hypothesize that optimistic bias will be more pronounced in conditions in which participants do not expect that their predictions will be put to the test and that their predictions will be less biased—and more accurate—in conditions in which they do. Studies 1 and 2 set the empirical groundwork for subsequent theoretical development and evaluation, establishing whether, in essence, the hypothesized differences in predictions and in prediction accuracy obtain and, thus, need to be explained. In Study 3, we next evaluated the necessity of critical feedback in the real versus hypothetical effects (cf. Sedikides et al., 2002; Shepperd et al., 1996). In Studies 3 and 4, we also tested a number of explanatory hypotheses suggested by work on temporal proximity (e.g., preemptive self-criticism, Gilovich et al., 1993; mood as information, Savitsky, Medvec, Charlton, & Gilovich, 1998; Shepperd et al., 2005) as well as a number of previously untested hypotheses derived from Trope and Liberman's (2003) construal level theory. We provide further detail about these hypotheses in the context of the studies designed to test them.

A Note on Accuracy

The present studies share a concern with evaluating the accuracy of individual predictions. Optimistic biases have been most commonly investigated with *relative appraisals,* that is, assessments of the extent to which people see themselves as more likely than others to experience positive events and less likely than others to experience negative events (e.g., Perloff & Fetzer, 1986; Weinstein, 1980). There are, however, several drawbacks to using relative appraisals as indicators of prediction accuracy. First, these measures reveal information about bias only in the aggregate and cannot, on their own, reveal whether any one prediction is accurate (e.g., if a majority within a group expects to perform better than the average performing member of that group, many of them will be right). Second, several recent lines of research suggest that the appearance of optimistic bias in relative appraisals may reveal more about respondents' egocentrism than about optimism or pessimism per se. For example, whereas people appear to be overly optimistic when stating relative chances for success on easy tasks, they appear to be overly pessimistic when stating relative chances for success on difficult tasks (Chambers & Windschitl, 2004; Kruger, 1999; Kruger & Burrus, 2004). Other anomalies—such as the observation that optimistic biases appear to be reduced in Eastern cultures when assessed by relative appraisal methods (e.g., Heine & Lehman, 1995) but robust when predictions are compared with obtained outcomes (e.g., Buehler, Griffin, Otsubo, Lehman, & Heine, 2000; see Buehler et al., 2002, for discussion)—further suggest that measures of relative appraisals may be assessing something more (or something less) than people's tendency to be optimistic.

In the context of evaluating potential moderators of optimistic biases, the ability to assess individual prediction accuracy is critical. In Gilovich et al.'s (1993) studies of the effects of temporal proximity, for example, it is certainly possible that the observed reduction in the better-than-average effect reflected greater accuracy among individual participants in the temporally proximate conditions (an interpretation we are inclined to favor), but it is also possible that these participants could have become mindlessly middling, cautiously responding near the midpoint of the available

response scale without becoming any more accurate in their predictions, or that some participants may have become overly pessimistic while others remained excessively optimistic (thus appearing average only in aggregate). A reduction in bias, therefore, does not necessarily imply a reduction in error (see, e.g., Buehler et al., 1994).

In the studies presented here, we assessed predictions in a variety of ways, including relative appraisals, but in Studies 1 through 3 we also obtained an unambiguous criterion (e.g., test performance) against which prediction accuracy could be assessed. With this criterion, deviations from accuracy could be evaluated in a number of useful but different ways. Evidence of *bias* was revealed by signed differences between predictions and performance. Evidence of prediction *error* was revealed by the extent of absolute deviation between prediction and performance. As an additional measure of prediction accuracy, we also assessed the correlation between predictions and performance.

Study 1: The Accidental Scavenger

In Study 1, participants were asked to estimate how well they would perform on a novel and ultimately highly involving activity: a scavenger hunt. Half of the participants were informed at the outset that they would be asked to complete the hunt; the remaining participants were not informed that they would complete the hunt until after they had made performance predictions. To provide a basis for comparing predictions with performance (i.e., a criterion for accuracy), all participants were next asked to complete the scavenger hunt, and their performance was recorded. We hypothesized that participants who did not expect to complete the scavenger hunt would exhibit excessive optimism, as is so commonly demonstrated in studies of people's personal forecasts, but that participants who were led to expect that they would complete the hunt would make more modest—and more accurate—predictions.

Method

Participants

Participants were 38 university undergraduates (58% male) who participated in exchange for research participation credit. One additional participant failed to respond to a majority of our dependent measures and was not included in the final sample.

Procedure

In order to protect against social facilitation effects and outright cooperation on the scavenger hunt task, participants were tested individually.

Manipulating task-completion expectations. We manipulated participants' expectations of whether they would be asked to complete the scavenger hunt by altering the manner in which the scavenger hunt was introduced. Half of our participants were told, from the start, that their task as participants would be to evaluate a scavenger hunt and then to complete the hunt that they evaluated. Remaining participants were told that they would be asked to evaluate the scavenger hunt, but no mention of actual performance was made.

The scavenger hunt. The scavenger hunt task was similar to one used in prior research (for details, see Armor & Taylor, 2003). The hunt was described as a 30-min activity in which participating "scavengers" would be required to find as many items as they could from a list of 32 objects (e.g., a safety pin, a party invitation, a library book). All items could be found, with varying degrees of effort and ingenuity, on the university campus. A number of simple rules (e.g., one must hunt alone; no item may be stolen or purchased for the purpose of the hunt, etc.) were included to give all participants a common frame of reference.

Predictions and task evaluations. In order to bolster the task-expectation manipulation, the phrasing of the task assessment and prediction measures was altered to correspond to the experimental condition: Participants who were led to expect that they would complete the scavenger hunt were asked to evaluate the hunt and to make performance predictions as if it were an impending reality (e.g., "How well *will* you do on the scavenger hunt?"), whereas participants who were not led to expect that they would complete the scavenger hunt were asked to evaluate the hunt and to make predictions as if it were hypothetical (e.g., "how well *would* you do, if you were asked to complete the scavenger hunt?").

Performance predictions were obtained both in general terms, by asking participants how well they thought they would [will] do on the scavenger hunt, on a scale ranging from 1 (*extremely poorly*) to 7 (*extremely well*), and in very specific terms by asking participants how many of the 32 items listed they thought they would [will] find in the allotted time. In order to assess relative appraisals, participants were also asked to estimate how well the "typical student" at their university would perform on the scavenger hunt (as with the personal predictions, predictions of others were requested in both general and specific terms); these estimates could then be compared with self-estimates in order to derive relative evaluations. An additional specific assessment asked participants to estimate how long it would take them to find half of the items on the list (in minutes) if they did not have a time limit.

For exploratory purposes, we asked participants to make several additional general assessments, including: how much fun participants thought the scavenger hunt would [will] be, how much they thought they would [will] enjoy participating in it, how difficult they thought the scavenger hunt task would [will] be, and how good they were at this sort of game. Responses to all general assessment questions were made on 7-point scales.

Performance. Once participants completed the dependent measures, they were asked to complete the scavenger hunt. Participants were equipped with a stopwatch so that they could keep track of elapsed time and a shopping bag in which to carry scavenged items. They were reminded of the rules and of the 30-min time limit and were then prompted to begin at the experimenter's cue (the experimenter also started a stopwatch at this time in order to have an independent assessment of how long participants took to complete the hunt). At the end of the scavenger hunt, participants returned to a different experiment room and were greeted by a different experimenter who was blind to the participant's task-expectancy condition and performance predictions (but not to elapsed time). This experimenter administered a brief performance assessment questionnaire and tallied the number of items collected.

Results

Preliminary analyses revealed no interaction effects with participant gender in this or any subsequent study. Thus, this variable is not discussed further.

Performance Predictions

As hypothesized, the task-expectation manipulation had a considerable impact on performance predictions, with participants in the task-expected condition being consistently less optimistic than participants in the task-unexpected condition. As can be seen in Table 1, in which means and corresponding inferential statistics are presented, the effects of the task-expectancy manipulation tended to be more pronounced when the performance predictions were assessed with the use of specific, unambiguous (and therefore

Table 1
Performance Predictions as a Function of Whether Participants Expected to Complete the
Scavenger Hunt (Study 1)

Measure	Task completion expectation		
	Not expected	Expected	$F(1, 36)$
General/global predictions			
Predicted performance (self)	4.21	3.58	1.40
Predicted performance (others)	4.32	4.16	0.18
Self–other difference	−0.11	−0.58	1.16
Specific predictions			
Predicted number of items found (self)	16.68	10.32	12.58**
Predicted number of items found (others)	16.42	12.74	4.37*
Self–other difference	0.26	−2.42	4.17*
Time to find half of items on list	40.74	76.73	7.12*
Other measures			
Task difficulty	4.63	5.05	1.06
Personal efficacy	4.68	3.84	2.87†
Anticipated enjoyment[a]	3.97	4.00	0.00

[a] Responses to questions assessing anticipated fun and anticipated enjoyment were almost perfectly correlated, $r(38) = .97$, and therefore were averaged prior to analysis.
† $p < .10$.   * $p < .05$.   ** $p < .01$.

potentially verifiable) response options. Although responses to the more global, ambiguously defined prediction and task-assessment questions were in the hypothesized direction, these differences rarely approached statistical significance. By contrast, the effect of task expectancy on responses to the more specific, objectively scaled prediction measures were clear-cut: Participants who were led to expect that they would complete the hunt thought that they would find considerably fewer items ($p = .001$), that other scavengers would also find fewer items ($p = .044$), and that they would require more time to find half of the items on the list ($p = .011$) than did participants who did not expect to complete the scavenger hunt.

It is worth noting that neither group displayed the often demonstrated better-than-average effect. Indeed, participants in the task-expected condition thought that they would perform worse, on average, than the average participant. This apparent over-pessimism effect was nearly significant when comparing participants' predictions for self and other on the general measures of predicted performance, $t(18) = 2.00$, $p = .061$, and clearly significant when comparing the number of items participants expected that they and others would find, $t(18) = 2.60$, $p = .018$. Even those who did not expect to complete the hunt, who generally appeared to be more optimistic, did not expect to outperform the average participant, both $ts(18) < 1$.

Performance

Performance on the scavenger hunt was assessed by tallying the number of items each participant returned. Despite the fact that participants came to the task with markedly different performance predictions, participants in the two experimental conditions did not differ in terms of performance: Those in the task-unexpected condition returned with 10.8 items, on average, whereas those in the task-expected condition returned with 11.2 items, $t(36) < 1.0$, *ns*.[3]

Accuracy, Error, and Bias

The principal analyses of these data involved comparing predictions with performance. As hypothesized, evidence of prediction accuracy, error, and bias differed markedly as a function of participants' experimentally induced expectations of whether they would complete the scavenger hunt (see Table 2). Participants who did not expect to complete the scavenger hunt were considerably, and unrealistically, optimistic, expecting to find, on average, over 50% more items ($M = 16.7$) than they actually found ($M = 10.8$), $t(18) = 4.25$, $p < .001$. The magnitude of this optimistic bias was significantly reduced in the task-expected condition, $t(36) = 4.24$, $p < .001$. In fact, participants who did expect to complete the hunt expected to perform somewhat worse ($M = 10.3$) than they actually performed ($M = 11.2$), though this apparent over-pessimism effect was not significant, $t(18) = 1.09$, *ns*. Thus, participants in this study did overestimate how well they would perform, displaying excessive optimism, but only if they did not expect to complete the scavenger hunt.

Results from several additional analyses suggest that participants in the task-expected conditions were not only less biased but were also more accurate (see Table 2). For example, the standard deviation of the prediction–performance discrepancies was significantly smaller in the task-expected condition than in the task-

---

[3] A substantial proportion of participants (53%) failed to complete the scavenger hunt within the 30-min time limit, though over three quarters (76%) returned within 1 min of the deadline. Only three scavengers (8%) returned more than 5 min late (two from the task-unexpected condition, one from the task-expected condition). Although neither the tendency to return late nor the overall time spent on the scavenger hunt differed significantly between conditions, we did conduct additional analyses that adjusted the number of items returned to correct for advantages of coming in late (cf. Armor & Taylor, 2003). These analyses yielded results that were substantively similar to the uncorrected return numbers. Thus, we report analyses of raw performance scores; one exception is noted in the text.

Table 2
Accuracy, Error, and Bias in Performance Predictions as a Function of Whether Participants Expected to Complete the Scavenger Hunt (Study 1)

	Task completion expectation	
Comparison of predictions to performance	Not expected	Expected
Prediction–performance discrepancy (bias)		
M	5.84***	−0.84
(SD)	(5.99)	(3.37)
Absolute deviation (error)		
M	6.47***	2.42***
(SD)	(5.26)	(2.43)
Prediction–performance correlation		
Pearson's r	.44	.76***

*** $p < .001$ (significantly different from zero).

unexpected condition, as determined by a Levene test, $F = 7.61$, $p = .009$, indicating smaller average discrepancies in this condition. A more direct assessment of absolute error, calculated simply as the absolute value of the difference between predictions and performance, was significantly smaller for the task-expected group than it was for the task-unexpected group, $t(36) = 3.05$, $p = .005$.

Another way of looking at the accuracy of participants' predictions is to examine the correlation between predictions and performance. Among those who expected to complete the scavenger hunt, this correlation was substantial and significant ($r = .76$, $p < .001$). Among those who did not expect to complete the scavenger hunt, the correlation between predictions and performance was more modest ($r = .44$; $p = .059$), although even this seemingly respectable correlation appears to have been inflated by a few participants who took considerably longer than the time allowed to complete their hunts (see Footnote 3). If one controls for the amount of time participants took to complete the hunt, the resulting partial correlations between predictions and performance were .79 ($p < .001$) in the task-expected condition and .33 ($p = .19$) in the task-unexpected condition. These partial correlations are significantly different from one another, $z = 2.06$, $p = .039$.[4] It appears, then, that the predictions of participants in the task-unexpected condition were not only optimistically biased but also truly unrealistic, insofar as these predictions bore very little relation to subsequent performance, whereas predictions made by participants in the task-expected condition were considerably more accurate.

Discussion

The results of Study 1 are consistent with the hypothesis that optimistic biases are not invariant within persons, but situated within contexts, and that these biases are more strongly expressed in situations in which they are less likely to be challenged or tested (Armor & Taylor, 1998, 2002). The results of Study 1 are also broadly consistent with the results of Gilovich et al.'s (1993) studies of temporal proximity but show that (a) the reduction of optimistic biases can occur for real (as opposed to hypothetical) tasks just as for proximate (as opposed to distant) ones and that (b) this reduction in bias in not an artifact of the method of measurement (i.e., relative appraisals). In fact, the measure of relative appraisals used in Study 1 revealed a seemingly anomalous finding: Had relative appraisals been used as a sole indicator of our

participants' prediction accuracy, participants who were not expecting to complete the task would have appeared to have been more accurate, not less (they expected, on average, to perform about as well as the average participant), whereas participants who did expect to complete the task appeared to have been excessively pessimistic (having expected to perform worse than average) despite having quite accurately predicted their own performance. Although these results were not expected, they are not entirely unprecedented: Savitsky et al. (1998) obtained similar worse-than-average effects in conditions of temporal proximity, and expectations of average performance in conditions of temporal distance, in a study designed to closely replicate the procedures of Gilovich et al. (1993). The relationship between different indicators of prediction accuracy were examined further in Study 2.

The scavenger hunt task used in Study 1 has a number of advantages (e.g., participants found it to be novel and engrossing), but it also has a number of disadvantages that may limit the generalizability of the results that we obtained. First, although participants appeared to be quite engaged with the task (many returned at full sprint, sweaty and out of breath), the scavenger hunt was ultimately just a game, and some may have found it to be trivial or unimportant. Second, participants in the task-expected condition may have achieved greater accuracy by applying a stop rule, that is, by finishing their hunt once they found the number of items they predicted they would find. (Several pieces of data argue against this possibility, however: Participants in the task-expected condition did not finish their hunts earlier than participants in the task-unexpected condition, and they did not find fewer items despite making more modest predictions.) Third, not all participants finished the scavenger hunt on time, which may have influenced the results (although corrections described in Footnote 3 suggest that this was not the case). To address these potential limitations in Study 2, we used a performance task that was familiar, important, and less vulnerable to a stop-rule strategy, in

---

[4] If one simply excludes data from participants who returned more than 5 min late, the prediction–performance correlation in the task-unexpected condition drops to .35 ($p = .18$); excluding data from participants who returned more than 1 min late drops the correlation to .19 ($p = .53$). Removal of similarly tardy scavengers from the task-expected group leaves the correlations virtually unchanged ($rs = .76$ and .80, $ps < .001$).

addition to being a task for which the time limit could be more strictly enforced.

Study 2: Pop Quiz

In Study 2, participants were asked to evaluate (and then to complete) a specially crafted test of questions that were described as being drawn from future versions of the Graduate Record Examination (GRE; Educational Testing Service, 2005). To ensure that this test would be important and familiar to our study participants, we recruited a sample of individuals who were both expert in taking standardized academic tests and invested in doing well on them (Yale University undergraduates).

Method

Participants

Seventy-five undergraduates (49% women) who were enrolled in an introductory psychology course at Yale University participated in the study in exchange for research participation credit.

Procedure

Participants were tested individually and in small groups of no more than three. In order to heighten the perceived importance of the GRE test, participants were told that the study was part of a joint project between the university psychology department and the Educational Testing Service (ETS) and that the experimenters would be asking the participants to evaluate a set of questions that the ETS was considering for use in future versions of the GRE.

Manipulating task-completion expectations. As in Study 1, half of the participants were informed upfront that they would be asked to complete the set of GRE questions after answering an initial set of questions about it; remaining participants were told that they would be asked to respond to questions about the test, but no mention was made that anyone would actually be completing the test until after they had responded to the prediction measures. All test materials and prediction measures were phrased in condition-appropriate ways (e.g., "you will be asked. . ." vs. "test takers would be asked. . .").

The test. The test itself was a time-limited, 20-item test consisting of moderate to difficult questions taken from verbal sections of past GREs. Participants in each test-expectancy condition were randomly assigned to evaluate (and then to complete) one of three different tests, each consisting of only one type of GRE test question: analogies, antonyms, or sentence completions. All participants were provided with a thorough description of their particular test as well as some factual information that could, in principle, be used as a basis for making predictions. For example, participants were told that students nationwide have, on average, answered approximately 8 of the 20 questions correctly when answering these questions as part of the full GRE general test. Participants were also provided with a representative sample question of approximately the same average difficulty as the items that would appear on their test, as well as an official ETS-supplied explanation of the correct answer (these explanations were obtained from the ETS website, at http://www.gre.org/practice_test/takesc.html).

Prediction measures. As in Study 1, participants were asked to make a general prediction ("How well do you think you will [would] do on the sample test?"), to which they responded on a 7-point scale, and a specific prediction ("How many out of 20 questions do you think you will [would] answer correctly?"). In order to assess relative appraisals, participants were asked to make a single prediction about how well they would perform relative to other students on an 11-point percentile scale ranging from 0% (*not better than any other Yale student*) through 50% (*better than half*) to 100% (*better than all other Yale students*).

For exploratory purposes, we also asked participants to evaluate their confidence in the accuracy of their prediction, their belief in how well they generally do on these kinds of tests, and how diagnostic they thought the test would be in terms of determining their academic potential. Responses to these questions were assessed on 7-point scales, with higher numbers reflecting greater confidence, efficacy, and diagnosticity.

Taking the test. After providing their responses, all participants were asked to complete the 10-minute GRE test. Participants were then asked a number of follow-up questions, probed for suspicion, and debriefed.

Results and Discussion

Preliminary analyses revealed unexpected differences in participants' Scholastic Aptitude Test (SAT) Verbal scores across conditions (test-unexpected $M = 719$, test-expected $M = 743$), $t(73) = 1.90$, $p = .06$. Accordingly, all analyses used these scores as a covariate. Additional analyses did not reveal any interactions with gender or test type; all subsequent analyses therefore collapse across these variables.

Performance Predictions

As in Study 1, and as can be seen in Table 3, the test-expectancy manipulation consistently influenced the predictions participants

Table 3

Performance Predictions as a Function of Whether Participants Expected to Complete the Graduate Record Examination (Study 2)

Measure	Test completion expectation		$F(1, 72)$
	Not expected	Expected	
Prediction measures			
General prediction (how well?)	4.57	3.90	4.71*
Specific prediction (how many?)	11.91	10.19	4.09*
Relative appraisal (better than %)	50.52	42.13	4.23*
Other measures			
Personal efficacy (testing ability in general)	5.53	5.38	0.46
Confidence in prediction accuracy	4.06	3.92	0.66
Diagnosticity of test	3.19	3.01	0.54

Note. Means and inferential statistics have been adjusted for Scholastic Aptitude Test scores.
* $p < .05$.

made about their performance on the GRE test. Participants in the test-unexpected condition thought they would perform better ($p = .033$), that they would answer more questions correctly ($p = .047$), and that they would outperform a larger percentage of students at their university ($p = .043$), than did participants in the test-expected condition. The manipulation did not, however, influence people's beliefs about their general test-taking ability, their confidence in the accuracy of their predictions, or the extent to which they saw the test as diagnostic of their academic potential.

Again, as in Study 1, neither group of participants believed that they would perform better than average. Although Study 2 used a more direct measure of relative appraisals, the results were similar: On this measure, participants in the test-unexpected condition give the appearance of being reasonably accurate, expecting on average to perform about average ($M = 51$ st percentile), while participants in the test-expected condition expected to perform significantly worse than average ($M = 42$ nd percentile), $t(37) = 2.78, p = .004$.

Performance

As in Study 1, performance did not differ as a function of the test-expectancy manipulation. Although there was a trend for participants who expected to complete the test to perform a bit better ($M = 10.2$) than participants who did not expect to complete the test ($M = 9.5$), this difference was not significant, $F(1, 72) = 1.15$, *ns*.

Accuracy, Error, and Bias

As in Study 1, participants who expected to complete the test not only made more modest predictions than did participants who did not expect to complete the test, but they made more accurate predictions as well (see Table 4). Participants who did not expect to complete the test significantly overestimated how many questions they would answer correctly, $t(37) = 3.55, p = .001$, showing a clear optimistic bias, whereas participants in the test-expected condition did not, $t(38) < 1$, *ns*. This reduction in bias is significant, $F(1, 72) = 6.09, p = .016$. The average absolute error, again taken as the mean absolute value of the difference between

Table 4
Accuracy, Error, and Bias in Performance Predictions as a Function of Whether Participants Expected to Complete the Graduate Record Examination Test (Study 2)

Comparison of predictions to performance	Test completion expectation	
	Not expected	Expected
Prediction–performance discrepancy (bias)		
M	2.45**	0.03
(*SD*)	(4.20)	(4.20)
Absolute deviation (error)		
M	4.46***	2.71***
(*SD*)	(2.64)	(2.64)
Prediction–performance correlation		
Pearson's *r*	.05	.52***

Note. Means and inferential statistics have been adjusted for Scholastic Aptitude Test scores; adjusted standard deviations are listed in parentheses.
** $p < .01$.   *** $p < .001$ (significantly different from zero).

predictions and performance, was also significantly greater among those who were not expecting to complete the test than among those who were expecting to complete the test, $F(1, 72) = 8.03, p = .006$. The difference in the correlation between predictions and performance was similarly substantial: For those not expecting to take the test, the correlation between predictions and performance was negligible ($r = .05, p = .76$), but for participants who did expect to complete the test, this correlation was positive and strong ($r = .52, p < .001$). These correlations are significantly different from one another ($z = 2.18, p = .029$).

Study 3

Although Studies 1 and 2 used very different performance tasks, the results were the same: In both, predictions were unrealistically optimistic in the task-unexpected conditions but impressively accurate in the task-expected conditions. Notably, in both studies, the tasks under consideration were quite specific, and the amount of information provided to participants was held constant across conditions. Therefore, between-condition differences in predictions and prediction accuracy in these studies cannot be attributed to differences in the amount of information available about the task. The explanatory hypotheses evaluated in Study 3 suggest other possibilities.

Three Explanatory Hypotheses

Feedback

One important distinction that has been left unresolved in Studies 1 and 2 is whether the effects of the real versus hypothetical nature of the task were due to the effects of expecting to complete the task, as we have suggested, or whether our effects have been the consequence of differences in *expectations about the possibility of critical feedback*. In Studies 1 and 2, expectations about completing the test were potentially confounded with expectations about the availability of evaluative feedback. Although no mention of feedback was made in either study, participants in the task-expected conditions may have nonetheless worried about the possibility of receiving feedback and made less optimistic predictions either (a) in an effort to avoid disappointment of unmet expectations (self-protection; e.g., Shepperd et al., 1996; Taylor & Shepperd, 1998; see also Josephs, Larrick, Steele, & Nisbett, 1992; Larrick, 1993) or (b) in an effort to avoid looking foolish in the eyes of others (public accountability; e.g., Study 2 of Regan, Gosselink, Hubsch, & Ulsh, 1975; Sedikides et al., 2002; see also Lerner & Tetlock, 1999; Tetlock, 1992; Tetlock & Kim, 1987). In Study 3, we attempted to disambiguate the effects of test completion expectations from feedback expectations by manipulating these expectations independently.

Mood

Mood has been found to mediate the effects of temporal proximity (Savitsky et al., 1998) and feedback proximity (Shepperd et al., 1996, 2005) on performance predictions and post-task performance evaluations. In these studies, it is generally assumed that the proximity of performance (or of feedback) induces feelings of nervousness, which is then interpreted as a physiological signal that one will not, or did not, perform so well (see also Schwarz &

Clore, 1983, 2003). Although it was not clear why this hypothesis would predict greater accuracy rather than greater pessimism, we tested for possible effects of mood in Study 3 by assessing mood at the time participants made predictions.

Construal Level

According to construal level theory (Trope & Liberman, 2003), various manipulations of "psychological distance" will meaningfully influence the level of abstraction at which actions, events, and persons are mentally represented. Although most studies of construal level theory have focused on manipulations of temporal distance (e.g., Liberman, Sagristano, & Trope, 2002; Liberman & Trope, 1998; Nussbaum, Trope, & Liberman, 2003), Trope and Liberman (2003) have speculated that other manipulations, including the real versus hypothetical nature of events and actions, may effectively manipulate psychological distance as well.

In the context of the present studies, construal level theory suggests that hypothetical events will be represented in more abstract, high-level terms and that real, proximal events will be represented in more concrete, low-level terms. As Trope and Liberman (2003) described the distinction, "high-level construals are relatively simple, decontextualized representations that extract the gist from available information," whereas "low-level construals tend to be more concrete and include subordinate, contextual, and incidental features of events. . . low-level construals are thus richer and more detailed but less structured and parsimonious than high-level construals" (p. 405). At a higher, more abstract level, the GRE test may be thought of as a meaningful academic exercise on which participants will want to perform well. Such a representation could thus motivate people to be optimistic, as successful performance could be seen as symbolic of the test taker's intellectual potential. At a lower, more concrete level, by contrast, the GRE test may be thought of in terms of broken pencils, smudgy erasers, and impending time limits—not the kind of high-level terms that would inspire optimistic predictions. When thinking of the test at this lower level, then, participants in the test-expected conditions may be more mindful of a host of contextual factors (e.g., testing conditions, their own level of alertness, etc.) that may be peripheral to the "gist" of the GRE test but relevant to making an accurate prediction.

In an effort to assess these different levels of construal in Study 3, we asked participants to do two things: to recall very specific details about the GRE test, and to indicate, in essence, how meaningful they thought the test would be. To the extent that participants in the test-unexpected condition were representing the GRE test in abstract, high-level terms, they should be expected to have greater difficulty remembering specific details of the test; participants in the test-expected conditions, by contrast, were expected to form more concrete, lower-level representations of the test and thus were expected to remember more test details. The higher-level representation of the GRE test was also expected to be a more meaningful representation. We hypothesized that participants in the test-unexpected conditions would thus see more value in the test, and attach greater significance to their imagined performance on the test, in comparison with participants in the test-expected conditions who were expected to "dismiss the forest for the trees" and to deny the significance of the test and their test performance.

Method

Participants

Seventy-five undergraduates (37 women and 38 men) from an introductory psychology course were recruited to participate in exchange for research participation credit. Four additional participants did not attend to experimental instructions and thus were not included in the final sample.

Procedure

The test and cover story were identical to those used in Study 2: Participants were informed that they would be asked to evaluate (and, in some conditions, to complete) a test containing GRE questions that the ETS was considering for future examinations.

Manipulating performance and feedback expectancies. Two thirds of participants were assigned to one of two test-expected conditions in which they were told that they would be taking the test. Of these participants, half were assigned to a "feedback-expected" condition: These participants were told that they would be scoring their own test at the end of the study and thus would know how they performed on the test. The remaining participants in the test-expected conditions were assigned to a "no-feedback" condition: they were told that the ETS would not allow participants to review the answers to test questions, that even the experimenter was not informed of the test answers, and that the experimenter would not be allowed to see the participants' responses. To make this manipulation believable, participants in the no-feedback condition were told that they would be asked to seal their materials from the session, including their tests and answer sheets, in a stamped envelope addressed to the Educational Testing Service and to deposit these envelopes into a U.S. mail bin that had been placed in the test room.

The remaining third of the participants was assigned to a test-unexpected condition identical to that of Study 2: They were told that they would be asked to answer questions about the test, but no mention was made of participants actually having to complete the test.

Predictions and other measures. The prediction questions were the same as in Study 2, except that the efficacy question was not included. In addition, participants were asked to estimate their best and worst possible scores on the test.

In order to assess participants' representation of the test (i.e., as an indicator of construal level), participants were asked a series of questions about how meaningful they thought the test would be: They were asked how important it was for them to perform well on the test, how pleased they would be if they performed better than they expected, and how disappointed they would be if they performed worse than expected. The participants were also asked how diagnostic they thought the test would be and how important it was to make an accurate prediction.

Participants were next asked to indicate how they felt "at that moment" by completing a brief measure of current mood (consisting of 10 items taken from the Positive and Negative Affect Schedule [PANAS]; Watson, Clark, & Tellegen, 1988; sample items included *enthusiastic, happy, distressed* [reverse scored] and *nervous* [reverse scored]; $\alpha = .71$). In order to assess the extent to which participants had attended to test details (i.e., as a second measure of construal level), all participants were then asked to recall specific pieces of information from the test description. In a free-recall task, administered first, participants were asked to simply recall as much information as they could about the test (i.e., from the one-page information summary that they had read earlier in the study). A subsequent cued recall test asked participants to recall four specific details from the information page (e.g., the amount of time given for the test, the number of questions answered correctly in a nationwide sample of GRE test takers, etc.).

As a manipulation check, participants in the two feedback expectancy conditions were asked to recall whether they would receive their scores on the test (all remembered correctly). After providing their responses, all

participants completed the 10-min GRE test and then were asked to complete several additional measures not central to the present investigation.

Results and Discussion

Primary hypotheses were tested with pairs of orthogonal contrasts conducted in the context of a series of univariate analyses of covariance (ANCOVAs) in which participants' SAT Verbal scores were entered as the covariate. The first of these contrasts analyzed the test-expectancy effect by testing for differences between the test-unexpected group and the two test-expected groups. The second contrast tested the effects of the feedback expectancy manipulation within the two task-expected conditions.

Performance Predictions

As was the case in Studies 1 and 2, performance predictions were influenced by the real versus hypothetical nature of the GRE test. As can be seen in Table 5, in comparison with participants in the two test-expected conditions, participants who did not expect to complete the test tended to make more optimistic general predictions ($p = .094$) and made significantly more optimistic specific predictions ($p = .010$), expecting to answer over 15% more questions correctly than did participants in the test-expected conditions. Participants who did not expect to complete the test

also tended to imagine a better best-case score ($p = .086$) and made significantly less pessimistic worst-case score predictions ($p = .003$) than participants in the test-expected conditions. In contrast to Studies 1 and 2, participants' predictions about their relative standing on the GRE test were not influenced by the test-expectancy manipulation.

As can also be seen in Table 5, the feedback expectancy manipulation did not significantly affect any of the prediction measures, suggesting that it is the knowledge that one actually has to complete the task, and not the (normally associated) opportunity for feedback, that influences people's predictions.

Performance

Performance was once again not influenced by the test-expectancy manipulation, $t(71) = 1.16$, $p = .25$, and it was also not influenced by the feedback expectancy manipulation, $t(71) = 1.06$, $p = .29$. On average, participants answered 10.2 questions correctly.

Accuracy, Error, and Bias

Our indices of prediction accuracy were strongly and significantly influenced by the test-expectancy manipulation but not by the feedback-expectancy manipulation (see Table 5). Although

Table 5

Effects of Test and Feedback Expectancy on Predictions, Prediction Accuracy, and Other Measures (Study 3)

	Experimental condition			Planned contrasts $t(71)$	
Measure	Test unexpected	No feedback expected	Feedback expected	Test expectancy	Feedback expectancy
Prediction Measures					
General prediction	5.19	4.87	4.60	1.70†	0.86
Specific prediction	13.99	12.21	11.93	2.66**	0.32
Best possible score	17.76	16.79	16.54	1.74†	0.33
Worst possible score	8.73	6.86	6.09	3.02**	0.86
Relative appraisal	51.99%	51.59%	46.83%	0.68	1.00
Accuracy, error, and bias					
Prediction–performance discrepancy (bias)					
M	4.39***	2.22**	1.03	2.83**	0.35
(*SD*)	(4.37)	(2.98)	(4.29)		
Absolute deviation (error)					
M	5.31***	2.96***	3.61***	2.96**	0.79
(*SD*)	(3.22)	(2.55)	(2.54)		
Correlation					
Pearson's *r*	−.14	.57**	.45*	$z = 2.66$**	$z = 0.53$
Potential Mechanisms					
Mood	3.32	3.26	3.46	0.30	1.42
Meaningfulness					
Importance of doing well	4.62	3.92	3.62	2.01*	0.58
Pleased if score is better	6.01	5.40	5.55	2.24*	0.52
Disappointed if worse	4.67	3.74	4.07	2.33*	0.84
Diagnosticity of test	4.61	3.39	4.29	2.51*	2.50*
Importance of accuracy	3.77	3.52	3.35	0.82	0.36
Memory					
Free recall	3.91	5.01	5.64	2.53*	0.96
Cued recall	2.46	3.05	3.41	2.99**	1.17

Note. The prediction measure and potential mechanism values are means. Scholastic Aptitude Test Verbal scores have been covaried out of the above analyses.

† $p < .10$.   * $p < .05$.   ** $p < .01$.   *** $p < .001$.

participants in both of the test-expected conditions tended to overestimate how well they would perform, neither group was as optimistically biased (combined $M = 1.63$) as participants in the test-unexpected condition ($M = 4.39$), $p = .019$. Viewed another way, among participants in the test-expected conditions, 1 in 3 (36%) met or exceeded their predicted score, and almost all (86%) met or exceeded their worst-case prediction; by contrast, among participants in the test-unexpected condition, only one in five (20%) met or exceeded their predicted score, and just over half (52%) met or exceeded their worst-case prediction.

As in Studies 1 and 2, participants in the test-expected conditions were not only less biased but were also better calibrated than participants in the test-unexpected condition. The average absolute error between predictions and performance was significantly lower among those who were expecting to complete the test (combined $M = 3.29$) than among those who were not expecting to complete the test ($M = 5.31$), $p = .004$. Moreover, participants who expected to complete the test made predictions that were significantly correlated with their actual scores ($r = .50$, $p < .001$), but participants who did not expect to take the test did not ($r = -.14$, ns), and this difference in correlations was significant ($z = 2.66$, $p < .01$). As with the other indicators of prediction accuracy, neither absolute error nor the correlations between predicted and actual scores were influenced by the feedback expectancy manipulation.[5]

Possible Mechanisms

The lack of effects of the feedback manipulation renders the two feedback hypotheses (self-protection and public accountability) less plausible: If participants had been motivated to make more accurate predictions out of a concern that their predictions might be invalidated or out of a concern that others might view them critically, the feedback expectancy manipulation would have been expected to have had a significant effect on prediction and prediction accuracy, and it did not. Although it remains possible that a more powerful manipulation of feedback—perhaps with the addition of an evaluative component, or a clear public audience—may have had a more discernible effect (see, e.g., Regan et al., 1975; Sedikides et al., 2002), participants' responses to the post-performance questions described in Footnote 5 (and to the task diagnosticity question, described below) suggest that they had attended to the feedback manipulation and were affected by it.

Another mechanism that does not appear to account for the real versus hypothetical effect is mood. As can be seen in Table 5, neither the test-expectancy manipulation nor the feedback expectancy manipulation had effects on participants' reported mood (this was true regardless of whether we looked at the composite mood score, as reported in Table 5, separate indices of positive and negative affect, or responses to individual items). Differences in mood, therefore, do not appear to account for the observed differences in predictions and prediction accuracy.[6]

Participants did differ, however, in how meaningful they thought the test would be (see Table 5). As predicted by construal level theory, participants in the test-unexpected condition reported that it would be more important for them to do well on the test ($p = .049$), that they would be more emotionally affected if they over- or underperformed in relation to their prediction ($ps = .028$ and $.023$) and that the test would be more diagnostic of their true

ability ($p = .014$) than did participants in the test-expected conditions. The feedback expectancy manipulation, by contrast, only influenced the perceived diagnosticity of the test, revealing the unsurprising finding that participants who did not expect to receive feedback believed the test would be less diagnostic than those who did expect feedback ($p = .015$). This effect—the sole significant effect of the feedback manipulation on pre-performance measures—is important because it demonstrates that participants were mindful of the manipulation and responded to it in a reasonable way (see also Footnote 5) and suggests that the lack of effects of this manipulation on the prediction and accuracy measures were not the consequence of a failed manipulation.

As can be seen in Table 5, the test-expectancy manipulation also influenced how much participants remembered about the test before they took it. Again in keeping with the construal-level hypothesis, participants in the test-expected conditions were more successful in recalling significant test details than were participants in the test-unexpected condition, and this was true both for free recall ($p = .014$) and cued recall ($p = .004$). As was the case with the prediction and prediction accuracy measures, memory for test details was not influenced by the feedback expectancy manipulations.

To determine whether the measures of meaningfulness and memory for details mediated the effects of the test-expectancy manipulation on prediction accuracy, we entered composite indices of meaningfulness ($\alpha = .77$) and memory ($\alpha = .70$) into a pair of multiple mediator models (see Kenny, Kashy, & Bolger, 1998; Preacher & Hayes, 2005), one testing mediation of the test-expectancy effect on prediction error, the other testing mediation of the effect on prediction bias. We evaluated the mediation hypotheses using a bootstrap approach advocated by Preacher and Hayes (2005; for additional discussion of the value of bootstrapping over other techniques for assessing mediation, such as the Sobel test, see Shrout & Bolger, 2002). Results of these analyses

---

[5] We do not wish to claim that the feedback manipulation had no effects on optimism. The null effects of the feedback manipulation reported in the text were observed on prediction measures that had been obtained before the test was completed and thus well before the threat of feedback was imminent. Additional measures obtained after participants had completed the test revealed that participants in all conditions became pessimistic after they completed the test (reporting that they had answered, on average, 0.59 fewer questions correctly than they actually did answer correctly, $F(1, 71) = 4.27$, $p = .042$; cf. Shepperd et al., 1996; Heath & Jourden, 1997) but that this pessimism effect was more pronounced among those expecting feedback ($M = -1.10$) than among those who did not ($M = -0.46$) and those originally assigned to the test-unexpected condition ($M = -0.21$). In fact, when looked at within condition, this post-performance pessimism only approached significance in the feedback-expected condition, $t(24) = 1.93$, $p = .066$. Because these post-performance procedures and results are beyond the scope of this article, they are not discussed further (for additional details, see Sackett, 2002).

[6] Notably, mood was assessed after predictions. Therefore, it is possible that participants in the test-expected conditions may have (a) experienced a relatively negative mood prior to making predictions, (b) made more modest predictions in response to that mood, and then (c) experienced relief from that negative mood as a function of having made more modest predictions. In this way, our lack of mood effects does not necessarily rule out mood as a mechanism linking task completion expectations to prediction accuracy. We thank James Shepperd for suggesting this interpretation.

reveal that reported meaningfulness, but not memory for test details, mediated the effects of the test-expectancy manipulation on our index of prediction bias. Specifically, the direct effect of test-expectancy on bias was rendered nonsignificant when the mediators were included in the model, $t(70) = 1.58$, $p = .12$. Moreover, the indirect effect through meaningfulness was statistically significant: Using Preacher and Hayes's bootstrapping procedure, we obtained a point estimate for this effect of $-.18$, with a bias-adjusted and accelerated 95% confidence interval of $-.44$ to $-.01$ (we can conclude that meaningfulness is a statistically significant mediator because this confidence interval does not contain zero). The indirect effect through memory, however, was not significant: The point estimate for this effect was $-.16$, with a bias-adjusted and accelerated 95% confidence interval of $-.52$ to $.03$. Neither meaningfulness nor memory for test details were found to mediate the effects of the test-expectancy manipulation on our index of prediction error.

Alternate Explanations?

Although the meaningfulness and memory results are consistent with a construal-level interpretation, alternative explanations are possible. For example, the finding that participants in the test-expected conditions appeared to downplay the meaningfulness of the GRE test could be interpreted as a self-protective strategy (dismissing the significance of the test could be an effective strategy for minimizing the impact of receiving the low scores that they expected to receive) rather than a result of lower-level construals posited by temporal construal theory; however, the lack of effects of the feedback expectancy manipulation would seem to render this interpretation less plausible. A different explanation could be invoked to explain the memory data: It is possible that participants in the task-unexpected conditions may have remembered fewer details about the test because they simply did not think very carefully about tasks that they believed would remain hypothetical. However, neither self-protection nor this "insufficient processing" account can parsimoniously explain why participants in the test-unexpected condition saw the test as more meaningful *and* remembered fewer details about it than did participants in the test-expected conditions.

Study 4: Making the Hypothetical Seem Real

Our final experiment was designed with two goals in mind. First, we sought to determine whether the kinds of thought processes invoked by the real prospect of completing a task could be induced without requiring participants to actually complete the task. In other words, we wanted to know whether personal predictions could be made to be less optimistic even when the task under consideration is understood to be purely hypothetical. If real tasks are somehow more motivating or have more implications for self-evaluation or self-protection, and if these features serve as incentives to be accurate, then removing these features by making the task hypothetical in all conditions should eliminate between-condition differences in participants' predictions. If, on the other hand, the processes involved are largely imaginative, then one should be able to mimic the conditions of taking a real test without actually asking participants to complete the test. Therefore, in Study 4, we asked all participants to imagine taking a test that they

knew they would not be asked to complete. Half of the participants were assigned to a "context-unspecified" condition in which they were asked to imagine our hypothetical test in the same way we had asked participants to imagine the test in our prior test-unexpected conditions, that is, without any specificity about when or where they would complete the test. Remaining participants were assigned to a "here-and-now" condition in which they were asked to imagine not only that they would be taking the test but also that they would be taking the test at that very moment (i.e., at that exact time and in that location). The simple addition of these temporal–local constraints on participants' conceptualization of the test was expected to make the hypothetical seem real—and thereby to influence participants' construal level of the test and their predictions of how well they would perform on it—without requiring them to actually complete the test.

As a second goal, we sought to further disentangle the construal-level hypothesis (which focuses on the *content* of processing) from the insufficient processing alternative (which focuses on *quantity* of processing). We did this, in part, by asking participants to elaborate on the factors that might influence their performance on the GRE test either before or after they were asked to make predictions. This procedure was intended both as a manipulation of the extent to which participants elaborated on the test before making predictions (allowing a test of the causal effect of more thorough cognitive processing on optimistic predictions) as well as a method for assessing cognitive content both before and after participants made their predictions. If participants in our previous task-unexpected conditions had been unrealistically optimistic because they had simply thought less thoroughly about the tasks in question, then the following should hold true: (a) Participants in the context-unspecified condition should generate fewer factors as relevant to their performance (insofar as the context-unspecified conditions are intended to replicate the task-unexpected conditions of Studies 1 to 3) and (b) Encouraging participants to think more thoroughly about the factors influencing their performance before they make their predictions should lead these participants to be less excessively optimistic.

Method

Participants

One hundred and two university students (64% women) were recruited to participate in a brief study in exchange for a small incentive (a candy bar or a cold drink).

Procedure

Participants were asked to make predictions about how well they thought they would perform on a test of sample GRE questions. The test description provided to participants was identical to the ones used in Studies 2 and 3. The experimenter made it clear to all participants that they would not actually be taking the test, and the experiment itself was presented as an exercise in imagination.

Half of our participants were randomly assigned to a "context-unspecified" condition—a condition directly analogous to the test-unexpected conditions of Studies 2 and 3—in which they were asked to imagine taking the GRE test without any constraints on when or where they should imagine themselves taking the test. Remaining participants were assigned to a "here-and-now" condition in which they were asked to

imagine that they really would be taking the test and that they would be doing so in the immediate present (i.e., at that moment and in that location).

To provide an experimental test of the insufficient processing hypothesis, participants in each of the imagination conditions were randomly assigned to either a "prior elaboration" condition (which was intended to encourage effortful processing when making predictions) or to a "delayed elaboration" condition (which was not). In the prior elaboration conditions, participants were asked to "list as many factors as you can think of that might influence your performance on this verbal test" *before* they were asked to make their performance predictions. In the delayed elaboration condition, participants were not asked to list influential factors until *after* they had made their performance predictions.

As in the previous studies, participants were asked to make both general ("how well?") and specific ("how many?") predictions, as well as a number of other assessments (how difficult they thought the test would be, how capable they would be of concentrating, how much they would enjoy the test, and how motivated they would be to perform well).

Results and Discussion

Effects on Predictions

As hypothesized, participants in the "here-and-now" conditions, who had been asked merely to imagine taking the test in the immediate present, made more modest predictions and task assessments than did participants in the context-unspecified (or "pure hypothetical") conditions. In comparison, participants in the context-unspecified conditions expected to perform better on the test (Ms = 5.54 vs. 4.78), $F(1, 98) = 11.4$, $p < .001$, and to answer more questions correctly (Ms = 15.0 vs. 13.6), $F(1, 98) = 4.82$, $p = .03$. Notably, participants in both imagination conditions expected to answer more questions correctly than participants actually had answered correctly when taking the same test in Studies 2 ($M = 9.8$) and 3 ($M = 10.2$), suggesting that both groups were optimistically biased. Nonetheless, the significant between-condition difference in predicted scores reveals that participants in the here-and-now conditions were significantly closer to these expected values.

In comparison with participants in the context-unspecified conditions, participants in the here-and-now conditions also indicated that they would enjoy the test less (Ms = 3.06 vs. 3.88), $F(1, 98) = 6.22$, $p = .014$, that they would be less capable of concentrating (Ms = 3.60 vs. 5.25), $F(1, 97) = 32.7$, $p < .001$, and that they would be less motivated to perform well (Ms = 4.36 vs. 5.62), $F(1, 98) = 15.5$, $p < .001$. The only responses that were unaffected by the imagination manipulation were the assessments of task difficulty ($F < 1$, *ns*).

In contrast to the effects of the imagination manipulation, which significantly influenced responses on five of our six dependent measures, the cognitive elaboration manipulation did not consistently influence predictions or task assessments. The only trend to approach significance was the question about anticipated motivation, with participants in the prior elaboration condition stating that they would be less motivated ($M = 4.64$) than participants in the delayed elaboration condition ($M = 5.29$), $F(1, 98) = 3.71$, $p = .057$ (all other Fs < 1). The order manipulation also did not interact with the imagination condition to influence either predictions (Fs < 1) or the other assessment measures (Fs < 2.9, ps > .09). These results are thus inconsistent with the insufficient processing hypothesis: Inducing participants to think more thoroughly about the factors that might influence their performance did not have a significant effect on their predictions, either as a main effect or through selectively making seemingly "neglectful" participants in the context-unspecified condition more mindful.

Factors Listed: Quantitative Differences

The lack of significant effects of the cognitive elaboration manipulation does not appear to have been the consequence of an unsuccessful manipulation: Participants in the prior elaboration conditions listed more factors ($M = 8.46$) than did participants in the delayed elaboration conditions ($M = 6.90$), $F(1, 98) = 4.47$, $p = .037$, suggesting that participants in the prior elaboration conditions had thought more thoroughly about these factors than had participants in the delayed elaboration conditions. Thinking thoroughly, then, does not appear to be a sufficient remedy for overly optimistic predictions. A second strike against the insufficient processing hypothesis is that participants in the context unspecified conditions, who did make more optimistic predictions, did not appear to reason less thoroughly (and thus generate a shorter list of causal factors) than participants in here-and-now conditions; if anything, there was a trend in the reverse direction (Ms = 7.06 and 8.25, respectively), $F(1, 98) = 2.37$, $p = .13$.

Factors Listed: Qualitative Differences

Coding and analysis of the content of participants' factor lists provided additional insight into the thought processes involved in the here-and-now and context-unspecified conditions. To facilitate coding, each participant's factor list was first partitioned into single-factor units, and the resulting factors were then coded along three independent dimensions (described below). In order to assess rater reliability, a second rater coded factor lists from a random sample of 20 participants drawn equally from all conditions (24.6% of all factors generated). Interrater reliability was acceptable for all three categories (agreement > 83%; kappas > .69).

Attribution. Drawing from conventional distinctions within attribution theory (e.g., Kelley, 1967), causes were coded as making reference to the *self* (person attributions), to the *test* (stimulus attributions), or to the testing *environment* (circumstance attributions). In order to test predictions from construal-level theory, self-attributions were further subdivided into *stable self* characteristics (such as traits, general abilities, and test-taking experience) and *unstable self* characteristics (such as current mood, alertness, and level of concentration).

Overall, more than two thirds of the factors listed identified aspects of the self, and the majority of these aspects referred to unstable characteristics (including mood states, feelings of confidence, alertness, health, and ability to concentrate). However, in keeping with the predictions of construal level theory, participants in the context-unspecified conditions were more than twice as likely to make reference to stable aspects of themselves, such as to their vocabulary, intelligence, and affinity for standardized tests, than were participants in the here-and-now conditions (Ms = 18% and 8%, respectively), $F(1, 97) = 5.49$, $p = .021$. Participants in the context-unspecified conditions were also twice as likely to make reference to the only other stable factor (i.e., the test), though this effect (7% vs. 3%) was only marginally significant, $F(1, 97) = 3.45$, $p = .066$. Participants in the here-and-now conditions, by contrast, made correspondingly greater reference to the two less

stable factors: unstable self characteristics (59% vs. 53%) and the environment (29% vs. 21%). None of these effects were moderated by the order of the factor listing procedure (all interaction Fs < 2.0, all ps > .15).

Causal certainty. Some of the causal factors that participants listed were described in such a way as to indicate that participants were certain of their relevance to predicting test performance (e.g., "I generally do well on these kinds of tests"; "timed tests are always bad"), and other factors were described so as to indicate that participants were not sure of their relevance (e.g., "the weather, maybe"; "I might get bored"). To capture this causal certainty distinction, factors were coded as either *definite* or as *possible/uncertain.*

Participants in the context-unspecified conditions expressed greater uncertainty about whether the factors they listed would be relevant to their performance on the GRE test, at least in the delayed elaboration conditions. Analysis of the proportion of factors participants described as being of uncertain relevance to their performance revealed a main effect of the imagination manipulation, $F(1, 97) = 6.03$, $p = .016$, that was driven by a significant interaction between the imagination and elaboration manipulations, $F(1, 97) = 7.12$, $p = .009$. In the prior-elaboration conditions, participants in the context-unspecified condition were uncertain about roughly the same proportion of factors ($M = 30\%$) as were participants in the here-and-now conditions ($M = 32\%$). In the delayed elaboration conditions, by contrast, participants in the context-unspecified condition were uncertain about a much larger proportion of factors ($M = 47\%$) than were participants in the here-and-now condition ($M = 15\%$). Although the greater uncertainty in the context-unspecified conditions is not surprising—this uncertainty may explain why participants in these psychologically distant conditions need to rely on abstract, gist-based representations of the target event (i.e., because so many of the specifics remain unknown)—the reason behind the interaction effect is unclear.

Causal valence. Coders also determined the expected effect of each causal factor (or of the participants' standing on these factors) on test performance. Causal factors were coded as implying either a *positive* effect on test performance (e.g., "I know a lot of vocabulary") or a *negative* effect on test performance (e.g., "lack of preparation"). In some cases, participants listed causal factors without specifying the nature of the effect it may have (e.g., "my mood might influence my performance") or explicitly indicated uncertainty about the likely effect (e.g., "the test format is odd. . . I'm not sure whether that will be a good thing or a bad thing"); these factors were coded as *unspecified/unknown,* and reflected a different kind of uncertainty.

Participants in the here-and-now conditions identified proportionally more factors that would have negative implications for their performance ($M = 41\%$) than did participants in the context-unspecified conditions ($M = 21\%$), suggesting that these participants may have been thinking more self-critically (see Gilovich et al., 1993, Study 4; see also Tetlock, 1992; Tetlock & Kim, 1987; cf. Eyal, Liberman, Trope, & Walther, 2004). Neither group was particularly inclined to identify factors that would have clearly positive effects on their performance (Ms = 13% and 15%, respectively). However, in keeping with the results of the causal certainty codes, participants in the context-unspecified conditions listed factors that were more frequently coded as having unspec-

ified or unknown effects ($M = 64\%$) than factors listed by participants in the here-and-now conditions ($M = 45\%$), again suggesting that participants in the context-unspecified conditions were facing greater uncertainty when formulating their predictions. In contrast to the results with the causal certainty codes, however, these causal valence results were not moderated by the elaboration order manipulation.

Summary. The coded data help explain both the logic of construal level theory and the general implausibility of the insufficient processing hypothesis. Participants in the context-unspecified condition were asked to confront a prediction task facing many unknowns—what the testing conditions would be like, what their own mental and physical state would be like at the time of the test, and so on—whereas participants in the here-and-now condition were confronted with the same prediction task with many fewer unknowns. It thus does not appear that participants in the test-unexpected conditions were simply thinking less thoroughly about the determinants of their performance (though they do appear to have been thinking about them less self-critically; see also Gilovich et al., 1993). The added uncertainty inherent in hypothetical events—"when?" "where?" "what kind of state will I be in?"— thus creates a more complex prediction task, one that may not be solvable through the simple application of cognitive effort.

General Discussion

As early as LaPiere's (1934) analysis of attitude–behavior inconsistencies, psychologists have been aware that people's claims about their beliefs and behaviors will differ depending on the social context in which they find themselves. Although LaPiere's classic study may be more familiarly discussed in the context of attitudes and behavior (e.g., Kraus, 1995; Wicker, 1969), the questions LaPiere asked of his respondents were, in fact, behavioral predictions, and those predictions were asked anonymously and without the expectation that they would subsequently be put to test. Results of LaPiere's inquiries, and many others since, have consistently shown that under hypothetical, task-unexpected conditions, people tend to make predictions that are largely in line with societal expectations and aspirations of the day (see also Linn, 1965; Sherman, 1980; Weinstein, 1980; Woodzicka & LaFrance, 2001). In short, they tend to make predictions that are unrealistically optimistic.

Results from the four studies presented here revealed that people's claims about their own expected outcomes will also differ depending on the context in which their predictions are made. Participants who had been asked to make predictions about performance on tasks they thought were hypothetical made predictions that were unrealistically optimistic in the truest sense of the term: They overestimated how they would perform, often by large margins, and their predictions were, at best, only weakly correlated with their actual performance. By contrast, participants who had been asked to make predictions about their performance on tasks that they knew they would complete made predictions that were impressively accurate: Their predictions, on average, did not deviate from their average performance, their overall rate of error was reduced, and their predictions were strongly correlated with their performance. These results are consistent with Armor and Taylor's (1998, 2002) suggestion that optimistic biases are not invariant within persons but rather are situated within contexts:

These biases do not appear to be an inevitable part of the prediction process, individual forecasters do not appear to be indiscriminately optimistic, and the expression of these biases appears to be largely dependent on the psychological context in which predictions are made (for similar conclusions, see also Armor & Taylor, 2003; Sackett & Armor, 2005a; Shepperd et al., 2006).

The results of these studies are also broadly consistent with Gilovich et al.'s (1993; Savitsky et al., 1998) work on the effects of temporal distance on subjective confidence and with Trope and Liberman's (2003) suggestion that task hypotheticality and temporal proximity may be effectively analogous as manipulations of "psychological distance." The results of these studies go beyond those of Gilovich et al., however, by showing that predictions for real, proximal events are not only *less biased* than predictions for hypothetical events, but that they are *more accurate* as well. By measuring performance as a criterion for prediction accuracy, estimates of accuracy, error, and bias could be assessed and compared across conditions.

It is worth noting that the measures of relative appraisals—that is, the measures of how well participants thought they would perform in comparison with other people—yielded results that, in Studies 1 and 2 at least, were perfectly misleading about the accuracy of participants' predictions. Although responses to these measures revealed a consistent decline in optimism in the task-expected conditions, they also suggested that participants in the task-unexpected conditions were the ones who were accurate (having claimed that they would perform about the same as others on average) and that participants in the task-expected conditions were biased (having claimed that they would perform worse than others). Although these results were not expected, they are not unprecedented: Savitsky et al. (1998) obtained similar results in the context of a study of the effects of temporal distance. One possible explanation is that the performance tasks we used in these studies may have been seen as particularly difficult and that the hypothetical nature of the task may have led participants in the task-unexpected conditions to overcome their normal tendency to be pessimistic when contemplating difficult tasks (Kruger, 1999; see also Burson, Larrick, & Klayman, 2006; Chambers & Windschitl, 2004). Understanding these differences, and the general correspondence (or lack of correspondence) between various measures of prediction accuracy and bias will be an important challenge for future research.

Another interesting result is that, in the three studies in which behavior was measured, we found little evidence of predictions leading to self-fulfilling prophecy. An impressive body of research suggests that, in Sherman's (1980) words, optimistic errors of prediction can be "self erasing," with even overly optimistic forecasts helping to bring about the expected outcomes (see also Armor & Taylor, 2003; Cervone & Peake, 1986; Sherman, Skov, Hervitz, & Stock, 1981; for a review, see Armor & Taylor, 2002). However, other studies have shown that even successful manipulations of people's expectations do not always lead to changes in behavior (e.g., Wilson & LaFleur, 1995). In our experiments, participants in the task-unexpected conditions made predictions that were dramatically more optimistic than the predictions made by participants in the task-expected conditions, but this optimism did not carry over to influence performance. Understanding when and why predictions sometimes do and sometimes do not lead to self-fulfilling prophecy is another important question for future

research. One possible hypothesis is that predictions for real, proximate behavior (which were here found to be highly correlated with performance) may be sufficiently tied to people's behavior to be self-fulfilling but that predictions for hypothetical behaviors (which were not found to be correlated with performance) may be relatively untethered and therefore not self-fulfilling (cf. Oettingen & Mayer, 2002). In other words, the truly unrealistic predictions made by participants in our task-unexpected conditions may have been too far removed from reality to have influenced subsequent behavior.

A Construal-Level Interpretation

Studies 3 and 4 also provide some initial evidence about why participants' predictions were unrealistically optimistic in the task-unexpected conditions but realistic in the task-expected conditions. Although many factors are likely to contribute to unrealistic optimism (Armor & Taylor, 1998; Buehler et al., 2002; Sackett & Armor, 2005b) and to prediction accuracy (Osberg & Shrauger, 1986), the results of Studies 3 and 4 suggest that the level of abstraction at which people represent real and hypothetical actions and events may contribute to the degree of accuracy, error, and bias present in people's predictions. As predicted by construal-level theory (Trope & Liberman, 2003), participants in the task-unexpected conditions appeared to represent the hypothetical tasks in more abstract, high-level terms than did participants in the task-expected conditions: They saw the GRE test as more meaningful yet remembered fewer particular details about the test and, when listing factors that might influence their performance, tended to list more stable and fewer fleeting factors, which suggests that the participants were representing their performance on the task in more general, gist-based terms.

In some respects, the construal-level results, too, were anticipated by LaPiere (1934), who argued that a participant's response to a question is merely "a symbolic response to a symbolic situation" (p. 230), though construal level theory offers greater specificity about how these situations (and people's imagined responses to them) will be symbolized. Results of Study 4 suggest that participants in the context-unspecified conditions were not only representing the "symbolic situation" (i.e., the performance task) in more abstract, high-level terms, but that they were also imagining their "symbolic response" to that task (i.e., their behavior) in more abstract, higher level terms as well. Research on self-evaluation has shown that, when people are asked to evaluate themselves in increasingly abstract terms, they tend to evaluate themselves much more positively than if they had been asked to evaluate themselves in more concrete terms (e.g., Dunning, Meyerowitz, & Holzberg, 1989). An intriguing question for future research is whether people are more inclined to rely on these abstract (and presumably positive) self-assessments when making predictions for hypothetical or otherwise psychologically distant tasks (cf. Ehrlinger & Dunning, 2003).

Although construal level theory provides a reasonable account of the real versus hypothetical effects obtained in these studies, several questions remain. For example, it is not clear how construal level theory would explain why participants in the task-expected conditions appeared to be pessimistic on the relative appraisal measures (though these results may be more of a challenge for the interpretation of relative appraisal measures of opti-

mism than for the application of construal-level theory). The construal theory account would also be further strengthened by additional data showing that the level of construal mediates the effects of task hypotheticality on optimistic predictions.

How Unrealistic Is Unrealistic Optimism?

The results of the studies reported here suggest a pair of very different conclusions about the nature of unrealistic optimism. On one hand, the predictions made by participants in the task-unexpected conditions demonstrate that people's predictions can be truly unrealistic. These data would appear to offer further validation to a growing list of studies that have characterized people's predictions as unrealistically optimistic (Armor & Taylor, 1998, 2002) and, like data from studies of the planning fallacy (Buehler et al., 2002), show how inaccurate people's predictions can be when their predictions are compared to performance. On the other hand, the results of these experiments could be taken to suggest that the prevalence of unrealistic optimism may be overstated, and that the frequently documented expressions of excessive optimism may be partly enabled by the anonymous, consequence-free prediction environments in which optimistic forecasts are so commonly studied. Our studies show that despite a strong reputation for being optimistic, people appear to have a clear capacity to put optimistic biases aside and to be quite accurate when making predictions.

We do not wish to claim, however, that optimistic biases are therefore necessarily the artifacts of anonymity, or that these biases are "hothouse creations" that will not be expressed in more consequential settings. Although a majority of studies on optimistic biases have been conducted in contexts free of consequence, these biases have been documented in increasingly diverse settings, including those in which consequences for inaccurate predictions are more immediate and real (e.g., Camerer & Lovallo, 1999; Glare et al., 2003; Lim, 2001; for review, see Dunning, Heath, & Suls, 2004). Our central argument is that context matters. Examining optimistic biases across meaningfully different contexts can help establish the generalizability of these biases, but cross-context comparisons can also usefully inform theory about the underlying processes involved in making predictions that are accurate, erroneous, or biased.

We would also like to emphasize that, despite the contextually situated nature of unrealistic optimism, the predictions people make when thinking about hypothetical events may nonetheless be quite important. Although these "symbolic responses" appear to be much revised in situations in which people recognize that their predictions may be tested, these predictions may nonetheless influence behavior and outcomes in the situations in which they are expressed. People are often asked to make predictions about events or outcomes that are, at the time of prediction, essentially hypothetical. People may also rely on optimistic estimates of what they would have done in a particular situation when judging what another person actually did in that situation, which could lead to overly critical interpersonal assessments. Further, psychologically, these optimistic forecasts may be important because they symbolize a future that one hopes to achieve and, thus, may be a source of comfort, pride, or self-esteem.

Trains in the Distance

In his own melancholy way, Paul Simon evoked our central findings when he sang that "everybody loves the sound of a train in the distance" (Simon, 1983, track 7). The success of the song, and this lyric in particular, is that we can all conjure up an image of the distant rumblings of a rolling train and agree that the sound evokes pleasing feelings of intrigue, adventure, and forward purpose. Like Simon's doomed-to-be-disillusioned lovers, participants in our psychologically distant, task-unexpected conditions were found to romanticize about even the most mundane of tasks, spinning optimistic fantasies about their performance on scavenger hunts and GREs. Moreover, participants in these task-unexpected conditions appeared to make their predictions without considering that if they were standing on the railroad tracks of immediate action, they might see things quite differently.

References

Armor, D. A., & Taylor, S. E. (1998). Situated optimism: Specific outcome expectancies and self-regulation. *Advances in Experimental Social Psychology, 30,* 309–379.

Armor, D. A., & Taylor, S. E. (2002). When predictions fail: The dilemma of unrealistic optimism. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 334–347). New York: Cambridge University Press.

Armor, D. A., & Taylor, S. E. (2003). Mindset, prediction and performance: Self-regulation in deliberative and implemental frames of mind. *Personality and Social Psychology Bulletin, 29,* 86–95.

Berra, Y. (1998). *The Yogi book.* New York: Workman.

Buehler, R., & Griffin, D. (2003). Planning, personality, and prediction: The role of future focus in optimistic time predictions. *Organizational Behavior and Human Decision Processes, 92,* 80–90.

Buehler, R., Griffin, D., Otsubu, Y., Lehman, D., & Heine, S. (2000). *A cross-cultural comparison of the planning fallacy.* Unpublished manuscript, Wilfred Laurier University, Waterloo, Ontario.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67,* 366–381.

Buehler, R., Griffin, D., & Ross, M. (2002). Inside the planning fallacy: The causes and consequences of optimistic time predictions. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 250–270). New York: Cambridge University Press.

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

Calderon, T. G. (1993). Predictive properties of analysts' forecasts of corporate earnings. *The Mid-Atlantic Journal of Business, 29,* 41–58.

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

Cervone, D., & Peake, P. K. (1986). Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior. *Journal of Personality and Social Psychology, 50,* 492–501.

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

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

Dunning, D., Meyerowitz, J. A., & Holzberg, A. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving

assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082–1090.

Educational Testing Service. (2005). *Graduate Record Examinations information and registration bulletin, 2005–2006.* Princeton, NJ: Author.

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.

Eyal, T., Liberman, N., Trope, Y., & Walther, E. (2004). The pros and cons of temporally near and distant actions. *Journal of Personality and Social Psychology, 86,* 781–795.

Gilovich, T., Kerr, M., & Medvec, V. H. (1993). Effect of temporal perspective on subjective confidence. *Journal of Personality and Social Psychology, 64,* 552–560.

Glare, P., Virik, K., Jones, M., Hudson, M., Eychmuller, S., Simes, J., & Christakis, N. (2003). A systematic review of physicians' survival predictions in terminally ill cancer patients. *British Medical Journal, 327,* 195–201.

Heath, C., & Jourden, F. J. (1997). Illusion, disillusion, and the buffering effect of groups. *Organizational Behavior and Human Decision Processes, 69,* 103–116.

Heine, S. J., & Lehman, D. R. (1995). Cultural variation in unrealistic optimism: Does the West feel more invulnerable than the East? *Journal of Personality and Social Psychology, 68,* 595–607.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review, 5,* 74–95.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 719–731.

Josephs, R. A., Larrick, R. P., Steele, C. M., & Nisbett, R. E. (1992). Protecting the self from the negative consequences of risky decisions. *Journal of Personality and Social Psychology, 2,* 26–37.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 192–240). Lincoln: University of Nebraska Press.

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

Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin, 21,* 58–75.

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

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

LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces, 13,* 230–237.

Larrick, R. P. (1993). Motivational factors in decision theories: The role of self-protection. *Psychological Bulletin, 113,* 440–450.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125,* 255–275.

Liberman, N., Sagristano, M. D., & Trope, Y. (2002). The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology, 38,* 523–534.

Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology, 75,* 5–18.

Lim, T. (2001). Rationality and analysts' forecast bias. *Journal of Finance, 56,* 369–385.

Linn, L. S. (1965). Verbal attitudes and overt behavior: A study of racial discrimination. *Social Forces, 44,* 353–364.

Norem, J. K., & Cantor, N. (1986). Defensive pessimism: Harnessing anxiety as motivation. *Journal of Personality and Social Psychology, 51,* 1208–1217.

Nussbaum, S., Trope, Y., & Liberman, N. (2003). Creeping dispositionism: The temporal dynamics of behavior prediction. *Journal of Personality and Social Psychology, 84,* 485–497.

Oettingen, G., & Mayer, D. (2002). The motivating function of thinking about the future: Expectations versus fantasies. *Journal of Personality and Social Psychology, 83,* 1198–1212.

Orne, M. (1962). On the social psychology of the psychological experiment. *American Psychologist, 17,* 776–783.

Osberg, T. M., & Shrauger, J. S. (1986). Self-prediction: Exploring the parameters of accuracy. *Journal of Personality and Social Psychology, 51,* 1044–1057.

Perloff, L. S., & Fetzer, B. K. (1986). Self–other judgments and perceived vulnerability to victimization. *Journal of Personality and Social Psychology, 50,* 502–511.

Preacher, K. J., & Hayes, A. F. (2005). *Asymptotic and resampling strategies for assessing and comparing indirect effects in simple and multiple mediator models.* Unpublished manuscript, University of North Carolina, Chapel Hill.

Quadrel, M. J., Fischhoff, B., & Davis, W. (1993). Adolescent (in)vulnerability. *American Psychologist, 48,* 102–116.

Regan, J. W., Gosselink, H., Hubsch, J., & Ulsh, E. (1975). Do people have inflated views of their own ability? *Journal of Personality and Social Psychology, 31,* 295–301.

Sackett, A. M. (2002). *Optimism and accuracy in performance predictions: An experimental test of the self-protection hypothesis.* Unpublished master's thesis, Yale University.

Sackett, A. M., & Armor, D. A. (2005a). *Manipulating the reasons for optimism: Reversing bias by shifting consequences.* Unpublished manuscript, Yale University.

Sackett, A. M., & Armor, D. A. (2005b). *Reasons for optimism: Strategic bias or unwanted error?* Unpublished manuscript, Yale University.

Savitsky, K., Medvec, V. H., Charlton, A. E., & Gilovich, T. (1998). "What, me worry?" Arousal, misattribution and the effect of temporal distance on confidence. *Personality and Social Psychology Bulletin, 24,* 529–536.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the life orientation test. *Journal of Personality and Social Psychology, 67,* 1063–1078.

Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 26, pp. 123–162). San Diego, CA: Academic Press.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45,* 513–523.

Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. *Psychological Inquiry, 14,* 296–303.

Sedikides, C., Herbst, K. C., Hardin, D. P., & Dardis G. J. (2002). Accountability as a deterrent to self-enhancement: The search for mechanisms. *Journal of Personality and Social Psychology, 83,* 592–605.

Shepperd, J. A., Sweeny, K., & Carroll, P. J. (2006). Abandoning optimism in predictions about the future. In L. J. Sanna & E. Chang (Eds.), *Judgments over time: The interplay of thoughts, feelings, and behaviors* (pp. 13–33). New York: Oxford University Press.

Shepperd, J. A., Grace, J., Cole, L. J., & Klein, C. (2005). Anxiety and outcome predictions. *Personality and Social Psychology Bulletin, 31,* 267–275.

Shepperd, J. A., Ouellette, J. A., & Fernandez, J. K. (1996). Abandoning unrealistic optimism: Performance estimates and the temporal proximity

of self-relevant feedback. *Journal of Personality and Social Psychology, 70,* 844–855.

Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology, 39,* 211–221.

Sherman, S. J., Skov, R. B., Hervitz, E. F., & Stock, C. B. (1981). The effects of explaining hypothetical future events: From possibility to actuality and beyond. *Journal of Experimental Social Psychology, 17,* 142–158.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and non-experimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422–445.

Simon, P. (1983). Train in the distance. On *Hearts and bones* [CD]. New York: Warner Brothers. (1990)

Sniezek, J. A., Paese, P. W., & Switzer, F. S. (1990). The effect of choosing on confidence and choice. *Organizational Behavior and Human Decision Processes, 46,* 264–282.

Taylor, K. M., Shepperd, J. A. (1998). Bracing for the worst: Severity, testing, and feedback timing as moderators of the optimistic bias. *Personality and Social Psychology Bulletin, 24,* 915–926.

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 331–376). New York: Academic Press.

Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology, 52,* 700–709.

Tetlock, P. E., & Lerner, J. (1999). The social contingency model: Identifying empirical and normative boundary conditions on the error-and-bias portrait of human nature. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 571–585). New York: Guilford Press.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110,* 403–421.

Watson, D., Clark, L., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070.

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

Weinstein, N. D. (1987). Unrealistic optimism about susceptibility to health problems: Conclusions from a community-wide sample. *Journal of Behavioral Medicine, 10,* 481–500.

Weinstein, N. D. (1998). *References on optimistic biases about risk, unrealistic optimism, and perceived invulnerability.* Unpublished manuscript, Rutgers University.

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

Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues, 25,* 41–78.

Wilson, T. D., & LaFleur, S. J. (1995). Knowing what you'll do: Effects of analyzing reasons on self-prediction. *Journal of Personality and Social Psychology, 68,* 21–35.

Woodzicka, J. A., & LaFrance, M. (2001). Real versus imagined gender harassment. *Journal of Social Issues, 57,* 15–30.

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://watson.apa.org/notify/ and you will be notified by e-mail when issues of interest to you become available!