Psychological Bulletin
2013, Vol. 139, No. 3, 702–724

0033-2909/13/$12.00 DOI: 10.1037/a0029951

Sobering Up: A Quantitative Review of Temporal Declines in Expectations

Kate Sweeny
University of California, Riverside

Zlatan Krizan
Iowa State University

Although people's outlook on the future tends to be characterized by hope and optimism, over time this outlook often becomes more dire. We review multiple theoretical accounts of this tendency to "sober up" as feedback about outcomes draws near, and we explicate factors critical to promoting these temporal declines in expectations. We then meta-analytically test the impact of these factors on temporal shifts in people's expectations about self-relevant outcomes. The findings reveal a robust and ubiquitous tendency to lower one's expectations as the moment of feedback draws near and implicate multiple contributing processes (declining control, changing accountability pressures, construal level changes, and affect management concerns) as important for this shift. Furthermore, the results reveal important differences in the methodological approaches used to examine temporal shifts in predictions and suggest that timing of predictions relative to outcomes and feedback plays a critical role in the nature of the phenomenon. Overall, the analysis reveals an important exception to positive illusions about the future and suggests that a time-sensitive turn toward pessimism has adaptive functions.

Keywords: optimism, pessimism, expectations, predictions, bracing

> Pessimism is, in brief, playing the sure game. You cannot lose at it; you may gain. It is the only view of life in which you can never be disappointed. Having reckoned what to do in the worst possible circumstances, when better arise, as they may, life becomes child's play. —Hardy

People's tendency to change their expectations as "the moment of truth" draws near has been the target of over 2 decades of research, with the first published study on the topic appearing almost 40 years ago (Nisan, 1972). Since then, research has documented downward temporal shifts in predictions of starting salaries (Shepperd, Ouellette, & Fernandez, 1996), corporate earnings forecasts (Calderon, 1993), score estimates for various reasoning tests (Gilovich, Kerr, & Medvec, 1993; Sanna, 1999; Savitsky, Medvec, Charlton, & Gilovich, 1998), driving test expectations (McKenna & Myers, 1997), performance expectations in a mock job interview (Sweeny, Shepperd, & Carroll, 2009), and medical test expectations (K. M. Taylor & Shepperd, 1998).

Collectively, these studies suggest a tendency for people to lower their expectations as the point of feedback approaches. This time-triggered pessimism stands in stark contrast to strong optimistic tendencies typically characterizing people's outlooks (e.g., Krizan & Windschitl, 2007; Peterson, 2000; S. E. Taylor & Brown, 1988). Moreover, people's apparent willingness to reduce or even relinquish their optimistic outlook seems nearly unbelievable in light of optimism's affective, social, and behavioral benefits (Armor et al., 2008; Bandura, 1982; Oettingen & Gollwitzer, 2009; Pezzo, Pezzo, & Stone, 2006; S. E. Taylor & Brown, 1988; Tyler & Rosier, 2009).

However, the psychological and behavioral consequences of expectations about personal outcomes are complex (Oettingen & Gollwitzer, 2009; Olson, Roese, & Zanna, 1996; Sweeny & Shepperd, 2010). Prior to an event or performance, expectations can either undermine or motivate efforts to improve outcomes. For example, some people harness low expectations to motivate themselves to perform well, while others harness their high expectations to motivate goal achievement (Oettingen, Pak, & Schnetter, 2001; Spencer & Norem, 1996). In both cases, expectations are not merely a reflection of efforts to prepare for an upcoming performance but instead serve as an effective motivating force. Following a performance (but preceding feedback), lowering expectations can mitigate stress and anxiety about upcoming feedback (Shepperd, Grace, Cole, & Klein, 2005) and can reduce disappointment following such feedback (Krizan, Miller, & Johar, 2010; Shepperd & McNulty, 2002; Sweeny & Shepperd, 2010). Furthermore, these reactions can undermine or motivate intentions to perform better in the future, depending on whether outcomes exceed or fall short of expectations (Sweeny, Dillard, & Fox, 2011).

Despite numerous empirical demonstrations of temporal declines in expectations and their psychological consequences, no systematic examination of the nature and sources of these time-specific changes currently exists. The goal of our article is to rectify this omission. Hereafter, we refer to declines in expectations toward the moment of truth as *sobering up*, or for ease, simply *sobering*. Webster's New College Dictionary (Severynse, 1995) defines *sober* as, among more obvious definitions, "subdued; devoid of frivolity, exaggeration, or speculative imagina-

This article was published Online First September 17, 2012.

Kate Sweeny, Department of Psychology, University of California, Riverside; Zlatan Krizan, Department of Psychology, Iowa State University.

We thank Paul Windschitl for his feedback on an earlier version of this article and Doug Bonett for his assistance with the analyses.

Correspondence concerning this article should be addressed to Kate Sweeny, Department of Psychology, University of California, 900 University Avenue, Riverside, CA 92521, or to Zlatan Krizan, Department of Psychology, W112 Lagomarcino Hall, Iowa State University, Ames, IA 50011. E-mail: ksweeny@ucr.edu or zkrizan@iastate.edu

tion" (p. 1046). We have selected this term because it vividly captures temporal declines in lofty, high expectations while remaining mute with respect to the causes of these declines. We also intend this term to be agnostic with regard to a comparison between expectations and reality (i.e., one's eventual outcomes). Research on temporal declines in expectations has examined changes in expectations over time regardless of whether these changes reflect unrealistic optimism, realism, or unrealistic pessimism. To draw a statistical analogy, *sobering* captures a downward slope in expectations without comment on the intercept of the line. Although the definition of sober as "devoid of . . . exaggeration" implies that initial expectations were unrealistically high, we use the term more generally to indicate a shift from a relatively high initial prediction toward a more modest later prediction.

In this article, we first describe a variety of approaches by which researchers have examined sobering. We argue that understanding the differences between these approaches is critical to an adequate understanding of the phenomenon and its causal factors. Next, informed by these differences, we present distinct theoretical accounts of sobering suggested by prior empirical and theoretical work. Drawing from these accounts, we present hypotheses regarding moderators of sobering that anticipate when and why sobering is most likely to manifest. We then examine these hypotheses meta-analytically. Following the meta-analysis, we synthesize the main findings, identify critical questions that remain unanswered, and outline key avenues for further research.

Sobering Up Over Time

Researchers have used many diverse methods and measures to capture the phenomenon of shifts in predictions over time. Most notably, some studies examine shifts in predictions *prior* to a performance or a personally relevant event (e.g., a course exam, intelligence test, medical test, etc., hereafter simply referred to as a *performance*), whereas others examine shifts *following* a performance but prior to feedback (i.e., learning the outcome of the performance; Figure 1). Furthermore, some studies examine shifts in predictions over time within the same participants (employing a within-subjects design), whereas others compare the predictions of people who are distant from versus people who are near to a performance or performance feedback (employing a between-subjects design). These studies also vary in the setting of the study, the domain of the performance, and how they measure predictions. Any study that contrasts predictions made at two or more time points prior to performance feedback has the potential to capture temporal changes in expectations and is thus considered in our analysis of sobering over time. However, as we discuss shortly, the above mentioned differences are likely to have both methodological and theoretical consequences for understanding of this phenomenon, and these consequences have so far gone unaddressed.

Although this diversity of methods and measures presents a golden opportunity to compare the extent to which people shift their predictions under a variety of contextual and methodological conditions, more often these studies are lumped together with little attention to their differences. Perhaps the most critical variation within this literature, and one that the literature has failed to acknowledge or address, is between studies that examine predictions *before* a performance during preparation and studies that examine predictions *following* a performance as people simply await feedback. One classic (and fairly typical) study of sobering *prior* to a performance examined students' predictions of their exam scores on the first day of class and then again on the day of the exam (Gilovich et al., 1993). Other studies of this type take a between-subjects approach by, for example, comparing the score predictions of participants who believe they will take a quiz in several weeks to those who believe they will take it a few minutes later (Nisan, 1972). Studies that examine sobering *following* a performance might have students make exam predictions immediately following a course exam and then again moments before they learn their exam grade (e.g., Shepperd et al., 1996), or they might compare the performance predictions of participants who believe they will receive test feedback in several weeks to those who believe they will receive feedback immediately (e.g., Sweeny & Shepperd, 2007). We should note that both pre- and post-performance paradigms compare predictions when performance *feedback* is relatively far versus relatively near. For example, exam grade feedback is closer on the day of the test than on the first day of class, just as exam grade feedback is closer on the day grades are available than immediately following the exam. However, as we discuss next, the reasons for shifts in predictions before and after a performance are likely to differ.

In the remaining pages of the introduction we outline four theoretical accounts for temporal shifts in predictions (see Table 1). By incorporating recent findings and integrating relevant constructs, these theoretical accounts systematically build on previous theoretical contributions to this topic (Carroll, Sweeny, & Shepperd, 2006; Sweeny, Carroll, & Shepperd, 2006). We describe how these accounts may apply to shifts in predictions that occur either prior to or following a performance and identify their implications for moderators of sobering. We close the introduction with a summary of specific hypotheses regarding our meta-analytic findings.

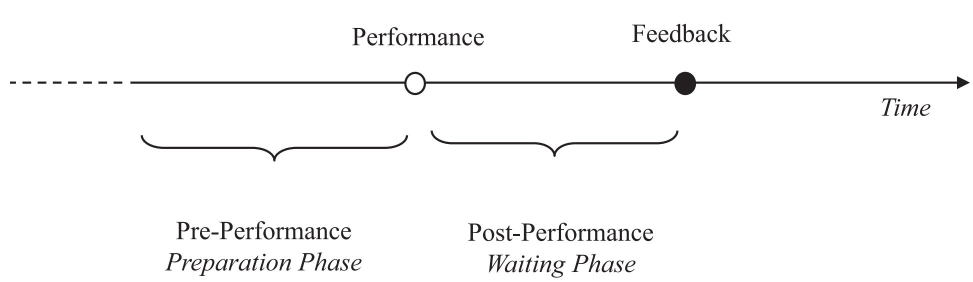

Figure 1. Temporal placement of pre- and post-performance estimates.

Table 1
Sources of Sobering During Preparation and Waiting Phases

	Preparation phase	Waiting phase
Declining control	⬂	
Construal level shifts	⬂	
Accountability pressures	⬂⬀	⬂⬀
Affect management	⬂	⬂

Note. Downward arrows indicate that the source is likely to magnify temporal declines in predictions. Upward arrows indicate that the source is likely to reduce temporal declines in predictions.

Theoretical Accounts for Sobering

There are numerous reasons why people might sober in their expectations over time. First, the extent to which people have control over performance outcomes declines over time, and with declines in control come declines in expectations (Shepperd et al., 1996). Research on the planning fallacy suggests that far in advance of a performance people typically are confident in their ability to take all appropriate preparatory actions and thus perform well, but this confidence declines as time for preparation runs out (Buehler, Griffin, & Ross, 1994). For example, people are very optimistic about completing their tax return on time far in advance of tax day, yet their optimism inevitably evaporates as the deadline approaches and their plans fail to materialize (Buehler, Griffin, & MacDonald, 1997). Once the performance is over, people change their focus from primary control (i.e., efforts to affect outcomes) to secondary control (i.e., efforts to change oneself and accept unchangeable outcomes; Morling & Evered, 2006; Rothbaum, Weisz, & Snyder, 1982) because their outcomes are already determined. Thus, declining control can help explain sobering prior to a performance, but reduced control is an unlikely explanation for sobering between the performance and feedback.

Second, construal level theory suggests that people may lower their predictions as a performance approaches due to a shift in how they think about the performance. People construe distant events in higher level, more abstract ways, whereas they construe proximal future events in lower level, more concrete ways (Trope & Liberman, 2003). For example, individuals construe more distant events in terms of broad, superordinate goals (e.g., "I broaden my horizons by reading a science fiction book") and more proximal events in terms of specific, subordinate goals (e.g., "I read the science fiction book by flipping pages"; Liberman & Trope, 1998; Vallacher & Wegner, 1989).

This difference in construal level has clear implications for the process by which people make performance predictions, as demonstrated by the finding that people base their distant performance predictions on abstract information and their proximal performance predictions on concrete information (Nussbaum, Liberman, & Trope, 2006). Thus, predictions of more distant events are often optimistic because they are based on goals and aspirations, whereas predictions of more proximal events are less optimistic because they are based on low-level details and feasibility considerations (Gollwitzer & Bayer, 1999; Trope & Liberman, 2003). Indeed, these differences in construal have direct consequences on people's performance predictions: When faced with an imminent performance, people's expectations become more pessimistic as they consider the many low-level contextual factors that might interfere with the performance (Armor & Sackett, 2006, Study 4). As with declining control, shifts in construal level are likely to contribute to changes in predictions only prior to a performance (Nussbaum et al., 2006). Once the construed event has passed, people presumably have encountered all performance-relevant details.

Third, accountability pressures represent another set of reasons for sobering, in this case both before and after a performance. People face increasing pressure to make accurate performance predictions as time passes, and they face the looming possibility that any unrealistic optimism will be challenged by either unexpected difficulty during the performance or an unexpectedly poor outcome (Carroll et al., 2006; Lerner & Tetlock, 1999). Not only might people worry about looking foolish if they overestimate their outcomes, they are also more likely to engage in self-critical thinking that reduces biased processing and undermines unrealistic optimism (Lerner & Tetlock, 1999; Tetlock & Kim, 1987). For example, people provide lower and more realistic performance estimates when they anticipate that others will learn their performance results, an outcome driven by a more critical appraisal of one's upcoming performance (Krizan, Scherr, & Windschitl, 2011).

In addition, as accountability pressures increase, people may look to their feelings as a source of information about their likelihood of success or failure (Clore, Schwarz, & Conway, 1994; Forgas, 1995). This process of using affect as information can lead to declines in predictions when anxiety increases as the feedback impends (Carroll et al., 2006). Similarly, as anxiety rises people might more readily call to mind negative information about their performance, consistent with research on mood-congruent memory (Bower, 1981; E. Johnson & Tversky, 1983), and these biased memories could lead to declines in performance predictions.

The types of accountability pressure discussed thus far focus on the accuracy of people's predictions, such that people may become increasingly concerned about inaccuracy as the moment of truth draws near. However, when people make multiple performance predictions over time, as participants do in many studies on sobering, they may also feel pressure to remain consistent with earlier predictions, just as people feel accountable to be consistent with unchangeable decisions (Lerner & Tetlock, 1999). To the extent that people feel accountable to previous performance predictions, they may be *less* likely to change their expectations over time. In short, accountability pressures could strengthen or weaken temporal declines in expectations and could do so both before and after a performance (see Table 1).

A fourth and final reason people may shift their predictions both prior to and following a performance is affect management. People can cope with unpleasant emotions via predictions in two ways, through *anticipatory* and *anticipated* affect management. Anticipatory emotions are visceral reactions (e.g., fear, anxiety) experienced in the face of uncertainty or risk, whereas anticipated emotions are not current emotional reactions but rather emotions that people *expect* to experience at a later time (Loewenstein, Weber, Hsee, & Welch, 2001). In the context of sobering, people engage in *anticipatory* affect management when they attempt to mitigate current feelings of anxiety about their outcomes by lowering their expectations. This tactic is analogous to lowering one's aspirations in order to decrease current feelings of pressure, to avoid "choking" when performance expectations are high (Baumeister & Showers, 1986). Although the precise mechanism by which low expectations relieve current anxiety is unknown, one study found that people did *not* lower their predictions in anticipation of feedback when they could attribute their anxiety to another source (i.e., a caffeinated drink; Shepperd et al., 2005).

Also, people engage in *anticipated* affect management, in which they lower their expectations to avoid disappointment in the case of a bad outcome (Shepperd & McNulty, 2002; Sweeny & Shepperd, 2010). People experience disappointment when their outcomes fall short of expectations (van Dijk & van der Pligt, 1997), so people can lower their expectations in an effort to reduce the likelihood of experiencing future disappointment (Rothbaum et al., 1982; Sweeny & Shepperd, 2010). Although these affective motivations for sobering have been documented in studies that examine post-performance predictions (Shepperd et al., 2005; Shepperd & McNulty, 2002; Sweeny & Shepperd, 2010), it seems likely that people also manage their anticipatory and anticipated emotions prior to performances in an effort to minimize emotional responses that could interfere with performance. That said, our discussion of the theoretical accounts of sobering reveals important potential differences between pre- and post-performance paradigms, and these difference have thus far gone unaddressed. Thus, one key goal of our quantitative review is to compare sobering between the two types of studies to determine whether shifts in expectations are similar in magnitude before and after a performance.

Moderators of Sobering

In addition to examining the overall robustness of sobering over time, our meta-analysis also examined a number of moderators that are likely to constrain the effect. Although the theoretical accounts just described are closest to the "heart" of sobering, few if any studies have provided direct tests of the accounts. Thus, the theoretical accounts served as the initial basis for generating hypotheses regarding potential moderating variables that *do* appear in studies on sobering.

We identified a number of situational factors that may intensify declines in control, construal level shifts, changes in accountability pressure, or affect management concerns. In addition to examining performance timing (whether the predictions are made prior to or following the performance), these factors included prediction timing (the overall distance between the prediction and the feedback), outcome importance (self-relevance of the performance domain), outcome familiarity (experience with the performance domain), study context (field vs. laboratory), nature of the design (within- vs. between-subjects), and prediction type (degree vs. likelihood). In the following section, we first define the constraint imposed by each potential moderator within the context of studies of sobering. We then frame each constraint in reference to its potential relationship with the theoretical accounts of control, construal level, accountability, and/or affect management. Finally, we provide testable hypotheses regarding the nature of the constraint's impact on the strength of sobering (i.e., the magnitude of shifts in expectations over time).

Performance Timing

As mentioned earlier, some studies of sobering examine shifts in predictions prior to a performance (*preparation phase*), and others examine shifts following a performance but prior to feedback (*waiting phase*). Critically, people may shift their predictions prior to a performance due to declining control, changes in event construal, increasing accountability pressures, or as an affect management strategy. In contrast, the only plausible sources of waiting phase shifts are accountability pressure and affect management; after the performance passes, control over performance outcomes and construal of the performance are unlikely to change over time (see Table 1). Thus, all else being equal, we hypothesized that people will show greater declines in expectations in studies that examine prediction shifts prior to performance than in studies that examine shifts following a performance.

Prediction Timing

Although a criterion for study selection was a comparison between distant and proximal performance predictions, studies varied in prediction timing on three dimensions: (a) time to feedback at the distant measurement point, (b) time to feedback at the proximal measurement point, and (c) difference in time to feedback between distant and proximal measurement points. Regarding the distant measurement point, studies examined predictions as far from feedback as many months or years and as close as 20 min. In many studies participants in the "distant feedback" condition believed that they would never receive feedback on their performance. Regarding the proximal measurement point, studies examined predictions as close to feedback as mere seconds away and as far as many months or years. These variations in timing resulted in further variation in the difference between distant and proximal measurement points, with the difference ranging from a literally infinite difference in timing (when some participants believed they would never receive feedback) to a difference of only 10 min (see Table 2).

Critically, greater lags in prediction timing should correspond to greater declines in control and greater shifts in construal level prior to a performance (Nussbaum et al., 2005), and to greater increases in accountability pressure (Lerner & Tetlock, 1999) and affect management concerns (Shepperd et al., 2005) both prior to and following a performance. Therefore, although few studies have examined whether people show greater downward shifts in their predictions when measured over longer periods of time, we hypothesized that people shift their expectations more when the difference between the distant and proximal measurement points was greater, given the opportunity for a larger cumulative effect of the processes just mentioned. Furthermore, due to the particularly precipitous increase in anxiety in the moments prior to feedback (Shepperd et al., 1996), we also hypothesized people would report the steepest declines in expectations when the proximal measurement point was closer to feedback, *regardless* of the difference between the distant and proximal measurement points. In other words, we anticipate declines to be stronger when the final prediction is made closer to the performance.

It is worth noting that we could define temporal distance in two ways in the context of studies on sobering: distance from the performance or distance from feedback. In studies of the pre-performance preparation phase, temporal distance declines over

Table 2
Descriptive Statistics

	Overall	Preparation phase only	Waiting phase only	Within-subjects only	Between-subjects only
Mean time to feedback (in min)	113,582 (109,280)	118,448 (100,592)	106,931 (121,611)	125,737 (117,120)	100,355 (100,108)
Distant time point	38,153 (90,108)	64,337 (111,878)	2,363 (6,306)	72,279 (115,255)	1,012 (2,193)
Proximal time point					
Mean time to feedback (subjective)					
Distant time point	5.1 (1.3)	5.2 (0.9)	5.0 (1.8)	4.9 (1.5)	5.4 (1.1)
Proximal time point	1.9 (0.9)	2.0 (0.8)	1.8 (1.1)	2.0 (1.2)	1.9 (0.7)
Mean difference in time to feedback (in min)	75,431 (95,739)	54,111 (64,311)	104,568 (122,062)	53,457 (86,482)	99,344 (100,764)
Mean difference in time to feedback (subjective)	3.2 (1.5)	3.2 (1.0)	3.3 (1.9)	2.9 (1.6)	3.6 (1.2)
Mean outcome importance	5.1 (1.3)	5.1 (1.3)	5.1 (1.2)	5.5 (1.1)	4.7 (1.3)
Mean outcome familiarity	4.1 (1.3)	4.4 (1.4)	3.7 (1.1)	4.1 (1.5)	4.2 (1.1)
Number of samples by type					
Preparation phase	41			20	21
Waiting phase	30			17	13
Within-subjects	37	20	17		
Between-subjects	34	21	13		
Lab	48	24	16	21	27
Field	23	17	14	16	7
Frequency of measurement types					
Degree—All types	76	50	26	35	41
Degree—Point prediction	45	24	21	26	19
Degree—Likert-type	19	15	4	5	14
Degree—Rank	12	11	1	4	8
Likelihood—All types	18	12	6	10	8
Likelihood—Gain	12	12	0	8	4
Likelihood—Loss	6	0	6	2	4

Note. Numbers in parentheses represent standard deviations. Frequencies for measurement types include multiple measures within a single sample.

time by both definitions. However, distance from performance and distance from feedback change inversely over time in studies of the post-performance waiting phase. We focus on temporal distance from feedback in our analysis due to the paucity of evidence supporting the salience of temporal distance from performance during the waiting phase. Construal level theory might suggest that were people to focus on the performance rather than the feedback (and thus perceive temporal distance as increasing rather than decreasing) they would increasingly base predictions on abstract considerations and thus *increase* their expectations of their performance over time (Nussbaum et al., 2006). Although no study we know of has directly tested this possibility, no study to date has found consistently upward shifts in predictions in anticipation of feedback. Thus, we focus on time remaining prior to feedback rather than time since the completion of a performance.

Outcome Importance

Participants in studies of sobering have faced feedback as critical as the results of a test for a severe medical condition (K. M. Taylor & Shepperd, 1998) and as trivial as the results of a scavenger hunt (Armor & Sackett, 2006). One way outcome importance may influence sobering is through its impact on affect management motivations. That is, people are more likely to feel anxious about an uncertain outcome that is important to them because they are more likely to suffer if the outcome is bad, and thus they may be more likely to lower their predictions to mitigate their rising anxiety (K. M. Taylor & Shepperd, 1998). Of course, accountability pressures may also loom larger for important outcomes to the extent that people are more likely to discuss their outcome predictions with others, thus becoming more sensitive to potential costs of inaccurate or boastful predictions (Lerner & Tetlock, 1999).

However, greater outcome importance may also promote maintenance of an optimistic outlook. People have a tendency to be *more* optimistic about more desirable outcomes (Krizan & Windschitl, 2007), and highly desirable outcomes are by definition more consequential and important. People may be particularly likely to maintain lofty expectations (i.e., *not* sober) for important outcomes that remain under their control (as in the case of pre-performance predictions), given optimism's role in fostering self-efficacy and motivating goal achievement (Armor & Taylor, 1998; Bandura, 1982; Oettingen & Gollwitzer, 2009). However, the "desirability bias" also occurs when control is absent (Krizan & Windschitl, 2007; Windschitl, Smith, Rose, & Krizan, 2010), which means people are often overoptimistic about outcomes they desire but cannot control (as in the case of post-performance predictions). These findings suggest that people may be *less* likely to sober in their predictions for highly important outcomes. Indeed, several studies confirm that optimism about important outcomes is less responsive to information suggesting that one's expectations should be lowered (Krizan et al., 2010; Krizan & Sweeny, in press; Massey, Simmons, & Armor, 2011).

On the other hand, most studies of desirability biases employ small monetary incentives or present situations in which people stand only to gain or maintain a status quo (Krizan & Windschitl, 2007). These scenarios may have few anxiety-provoking consequences for the participants (Shepperd, Findley-Klein, Kwavnick, Walker, & Perez, 2000) and thus may be insufficient to trigger the strong affective reactions participants likely had in many of the studies considered in our review. Thus, even if more consequential outcomes encourage higher expectations (Krizan & Windschitl, 2007), expectations for outcomes that involve substantial personal losses may be more sensitive to temporal declines than expectations about outcomes that do not (K. M. Taylor & Shepperd, 1998). Therefore, we tested these competing hypotheses by examining whether people lower their expectations more or less for outcomes that are more personally consequential (i.e., that had greater potential gains or losses).

Outcome Familiarity

Participants in studies of sobering may be more or less familiar with the performance domain and thus may be able to predict their performance outcomes with more or less confidence and accuracy. Furthermore, the predictions of people who lack experience with the domain, and thus confidence in their predictive accuracy, may be more susceptible to accountability pressures as a performance or feedback draws near (Tetlock, Skitka, & Boettger, 1989). By this reasoning, we would expect that people will lower their expectations less when outcomes are familiar because people might more readily shift their predictions when they lack confidence in their accuracy. On the other hand, if people have experience with a particular performance domain or outcome, they may also have experience with rising anxiety in the face of feedback and with disappointment over unexpected bad outcomes and thus experience greater concerns over affect management. By this reasoning, we would expect that people would lower their expectations *more* when outcomes are familiar because people are also familiar with the benefits of lowering their expectations as feedback draws near. Thus, we examined the relationship between outcome familiarity and sobering in order to test these two competing hypotheses.

Study Context

Despite the stereotypical reliance on laboratory studies within psychology, the literature on sobering is replete with studies of naturally occurring performance contexts. In fact, almost one third of studies on sobering are field studies that examine students' predictions of their performance on a course exam or project (see Table 3). Although lab studies of sobering typically use paradigms that have high mundane and experimental realism (e.g., exam-type predictions, interview feedback, etc.), it is possible that these studies do not quite recreate the anxiety-provoking nature of "real life" performances and thus do not elicit affect management concerns with the same intensity. Thus, we tentatively hypothesized that people would show greater declines in expectations in field studies than in lab studies.

Nature of Study Design

As mentioned earlier, sobering has been examined in both within- and between-subjects designs. Within-subjects studies of sobering take one of two forms. Most often, within-subjects studies use a longitudinal approach, such that participants make performance predictions when a performance or performance feedback is relatively distant and then again when the performance or feedback is nearer. Alternatively, some studies manipulate prediction timing within subjects, such that participants make performance predictions first when they believe feedback to be distant (or nonexistent) and then later when they are (unexpectedly) told that feedback is imminent. Between-subjects studies of sobering compare the performance predictions of participants who believe that a performance or feedback is relatively distant or nonexistent with predictions of participants who believe that the performance or feedback is relatively imminent.

We suspect that participants' psychological experience differs in these two types of studies. Participants in between-subjects studies report only one performance prediction, whereas participants in within-subjects studies report multiple predictions and thus record different predictions within a single study if they shift their expectations over time. To the extent that participants feel accountable to their initial performance predictions, they may be unwilling to report a different prediction later in the study, even if their personal expectations have changed (Lerner & Tetlock, 1999). Of course, participants in some studies have reported different predictions over time even when they report multiple predictions on the same piece of paper (Shepperd et al., 1996), but nonetheless, we hypothesized that people would lower their expectations less in within-subjects designs due to a commitment to initial performance predictions.

Prediction Type

Finally, the form of the prediction may have important consequences for the extent of sobering over time. Most broadly, predictions fell into two categories: (a) *degree* of performance success, which included point predictions (e.g., a specific exam score), relative rankings among other participants, and Likert-type measures of success or confidence, and (b) *likelihood* of performance success, which included probability judgments of both loss (e.g., probability of failure) and gain (e.g., probability of success). We hypothesized that people would lower their expectations more when making degree predictions than likelihood predictions. We reasoned that people might more readily shift their predictions when the predictions are more falsifiable and thus more subject to accountability pressures and affect management concerns, and it is easier to falsify a prediction of degree than likelihood. To illustrate, a participant who predicts a score of 15 out of 20 on an exam will be proven wrong by any score other than a 15 out of 20. In contrast, a participant who predicts a 30% chance of failure is not proven wrong by a passing grade; the participant simply beat the odds (see Windschitl et al., 2010, for similar reasoning).

Summary

To summarize, we presented four theoretical accounts of temporal declines in expectations in the face of impending feedback, and based on these accounts, we constructed hypotheses regarding the impact of various contextual factors on the tendency to sober over time. First, we anticipated that declines in expectations during the waiting phase following a performance would be smaller than shifts during the preparation phase prior to performance, given the potential for more factors to contribute to the changes in expectations during the preparation phase. Second, we anticipated that larger time intervals would yield larger declines in expectations

Table 3
All Included Studies With Effect Size and Study Characteristics

Reference and condition	*d*	*n*	Performance timing	Prediction timing: Distant point	Prediction timing: Near point	Nature of study design	Study context	Performance/event	Prediction type
Armor & Sackett (2006)									
Study 1	0.39	38	Preparation	Never	30 min	Between	Lab	Scavenger hunt	L–t
Study 2	0.50	75	Preparation	Never	<60 min	Between	Lab	GRE-style items	L–t, S
Study 3	0.18	50	Preparation	Never	<60 min	Between	Lab	GRE-style items	L–t, S
Carroll et al. (2009)									
Study 1									
No threat	−0.06	16	Preparation	Months or years	Months or years	Within	Lab	Admission to MA program	Lkhd
Unspecified threat	−0.12	16	Preparation	Months or years	Months or years	Within	Lab	Admission to MA program	Lkhd
Partly specified threat	−0.04	16	Preparation	Months or years	Months or years	Within	Lab	Admission to MA program	Lkhd
Fully specified threat	0.75	16	Preparation	Months or years	Months or years	Within	Lab	Admission to MA program	Lkhd
Study 2									
No threat	−0.06	17	Preparation	Months or years	Months or years	Within	Lab	Admission to MA program	Lkhd
Unspecified threat	−0.11	17	Preparation	Months or years	Months or years	Within	Lab	Admission to MA program	Lkhd
Partly specified threat	0.02	17	Preparation	Months or years	Months or years	Within	Lab	Admission to MA program	Lkhd
Fully specified threat	0.95	17	Preparation	Months or years	Months or years	Within	Lab	Admission to MA program	Lkhd
Gilovich et al. (1993)									
Study 1									
Psych students	1.17	54	Preparation	1–3 months	1–3 days	Within	Field	Course exam	S
Law students	0.31	28	Preparation	1–3 months	1–3 days	Within	Field	Course exam	S
Study 2	0.97	51	Preparation	1–3 months	Minutes	Within	Lab	Various tasks	R, L–t
Study 4	1.10	40	Preparation	1–3 months	Minutes	Between	Lab	Memory and persuasion tasks	R, L–t, $
Jiga-Boy (2008)									
Study 3	0.91	33	Preparation	8 week	2–3 days	Within	Field	Course exam	R, S
Study 4a	0.86	11	Preparation	7 weeks	2–3 days	Within	Field	Course exam	R, S
Study 4b	0.43	20	Preparation	7 weeks	2–3 days	Within	Field	Course exam	R, S
Kettle & Häubl (2010)	0.17	271	Preparation	9–17 days	0–9 days	Between	Field	In-class presentation	R
Krizan (2008)									
Study 1	0.37	107	Waiting	Never	Minutes	Between	Lab	GRE-style items	R, S
Monga & Houston (2006)									
Study 1 (chosen product)	0.56	22	Waiting	2 days	<1 day	Within	Lab	Camera performance	L–t
Study 2	0.76	50	Waiting	30 days	Immediate	Between	Lab	Camera performance	L–t
Nisan (1972)									
Failure orientation	0.13	50	Preparation	30 days	Minutes	Between	Lab	Reasoning test	Lkhd
Success orientation	1.51	50	Preparation	30 days	Minutes	Between	Lab	Reasoning test	Lkhd
Nussbaum et al. (2006)									
Study 2									
Easy items	0.30	36	Preparation	4 weeks	5–10 min	Between	Lab	Trivia quiz	Item value
Hard items	1.55	36	Preparation	4 weeks	5–10 min	Between	Lab	Trivia quiz	Item value
Study 3	0.60	62	Preparation	4 weeks	5–10 min	Between	Lab	Trivia quiz	Item value
Study 4									
Easy items	0.06	41	Preparation	4 weeks	5–10 min	Between	Lab	Trivia quiz	Lkhd
Hard items	0.89	40	Preparation	4 weeks	5–10 min	Between	Lab	Trivia quiz	Lkhd
Pennington & Roese (2003)									
Study 1	0.68	80	Preparation	2–3 weeks	2–3 days	Within	Field	Course exam	L–t, S
Regan et al. (1975)									
Study 2	0.50	19	Waiting	Never	5 min	Between	Lab	Interview	L–t
Sanna (1999)									
Study 1	0.37	36	Preparation	25 days	2–3 days	Within	Field	Course exam	L–t
Study 2	0.75	73	Preparation	30 days	2–3 days	Between	Field	Course exam	L–t

Table 3 (*continued*)

Reference and condition	*d*	*n*	Performance timing	Prediction timing: Distant point	Prediction timing: Near point	Nature of study design	Study context	Performance/event	Prediction type
Study 3									
Bad mood	0.43	30	Preparation	1–2 months	<1 hour	Between	Lab	Memory and anagram tasks	R, L–t
Neutral mood	1.46	30	Preparation	1–2 months	<1 hour	Between	Lab	Memory and anagram tasks	R, L–t
Sanna & Schwarz (2004)									
Control	0.30	30	Preparation	30 days	2–3 days	Between	Field	Course exam	L–t, S
3 failure thoughts	−0.06	30	Preparation	30 days	2–3 days	Between	Field	Course exam	L–t, S
12 failure thoughts	0.07	30	Preparation	30 days	2–3 days	Between	Field	Course exam	L-t, S
3 success thoughts	0.11	30	Preparation	30 days	2–3 days	Between	Field	Course exam	L–t, S
12 success thoughts	0.10	30	Preparation	30 days	2–3 days	Between	Field	Course exam	L–t, S
Savitsky et al. (1998)									
Study 1	0.65	88	Preparation	30 days	<1 hour	Between	Lab	Memory and anagram tasks	R, L–t, $
Shepperd (2010)	0.26	21	Waiting	Never	Immediate	Within	Lab	Breast cancer risk	Lkhd
Shepperd et al. (1996)									
Study 1									
Sophomores	0.11	31	Preparation	2.5 years	2 years	Within	Field	Starting salary	$
Juniors	0.12	22	Preparation	1.5 years	1 year	Within	Field	Starting salary	$
Seniors	0.31	29	Waiting	6 months	2 weeks	Within	Field	Starting salary	$
Study 2	0.16	144	Waiting	50 min	Immediate	Within	Field	Course exam	S
Study 3									
Low self-esteem	0.09	36	Waiting	50 min	Immediate	Within	Field	Course exam	S
High self-esteem	0.04	35	Waiting	50 min	Immediate	Within	Field	Course exam	S
Shepperd et al. (2005)									
Misattribution	0.09	51	Waiting	3 days	15 min	Between	Lab	Reasoning task	S
No misattribution	0.64	50	Waiting	3 days	15 min	Between	Lab	Reasoning task	S
Sweeny (2010a)									
Study 1	0.15	83	Waiting	5 days	5 min	Between	Lab	Reasoning task	S, Lkhd
Study 2	0.29	83	Waiting	30 days	7 days	Between	Lab	Unexpected bill	Lkhd
Study 3	0.33	161	Waiting	4–6 weeks	5 min	Between	Lab	Rating of personal photo	S
Sweeny (2010b)									
Study 1	0.47	80	Waiting	4–6 weeks	5 min	Between	Lab	Rating of personal photo	S
Study 2	0.10	113	Waiting	Never	10 min	Between	Lab	Rating of personal video	L–t
Sweeny (2010c)	0.40	108	Waiting	Never	5 min	Between	Lab	Reasoning task	S
Sweeny & Dillard (2010)	0.04	60	Waiting	50 min	Immediate	Within	Field	Course exam	S
Sweeny & Shepperd (2007)									
Study 1									
Low base rate	0.78	42	Waiting	Never	5 min	Between	Lab	Reasoning task	Lkhd
High base rate	0.57	35	Waiting	Never	5 min	Between	Lab	Reasoning task	Lkhd
Sweeny et al. (2009)									
Study 1	0.15	25	Waiting	Never	5 min	Within	Lab	Interview performance	S
Study 2	0.39	68	Waiting	Never	5 min	Within	Lab	Interview performance	S
Taylor & Shepperd (1998)									
Severe disease	0.22	17	Waiting	21–28 days	<1 hour	Within	Lab	Test for (bogus) disease	Lkhd
Mild disease	−0.07	17	Waiting	21–28 days	<1 hour	Within	Lab	Test for (bogus) disease	Lkhd
Terry & Shepperd (2004)									
Study 1	0.25	39	Waiting	Never	Immediate	Within	Lab	Social ratings	S
Study 2 (feedback)	0.31	27	Waiting	20 min	Immediate	Within	Lab	Social ratings	S
van Dijk et al. (2003)									
Relevant/immediate	0.42	40	Waiting	30 min	Immediate	Within	Lab	IQ test	S
Relevant/delayed	0.03	40	Waiting	2 weeks	2 weeks	Within	Lab	IQ test	S
Irrelevant/immediate	0.09	40	Waiting	30 min	Immediate	Within	Lab	IQ test	S
Irrelevant/delayed	0.02	40	Waiting	2 weeks	2 weeks	Within	Lab	IQ test	S
Wicker et al. (2004)									
Sample 1	1.06	41	Waiting	4 weeks	2–3 days	Within	Field	Course exam	L–t, S
Sample 3	0.33	68	Waiting	4 weeks	2–3 days	Within	Field	Course exam	L–t, S

Note. Performance timing is either preparation phase or waiting phase. Prediction timing refers to the time until feedback. For prediction type, S = score; L-t = Likert-type; R = rank; Lkhd = likelihood; and $ = monetary estimate; MA = master of arts; GRE = Graduate Record Exam.

and that these declines would be especially strong when proximal predictions are solicited very close to feedback. Third, we proposed contrasting hypotheses regarding the effect of outcome importance and familiarity on these temporal changes. Fourth, we anticipated that within-subject designs and predictions of likelihood may yield weaker declines due to increased commitment to prior predictions and their lower falsifiability, respectively. Of these hypotheses, only the role of outcome importance has received direct empirical attention. As such, our meta-analysis provides unique insight into critical theoretical and methodological considerations for research on sobering. With these hypotheses in mind, we next describe how we identified relevant investigations and outline the methodological approach we used to aggregate and compare studies along dimensions reflected in these hypotheses.

Method

In this section we present information about study inclusion criteria and methods of literature search, and we specify procedures we followed for study retrieval, data extraction, and coding of relevant variables necessary to conduct the review (see Cooper, 1998). We also describe the statistical methods we employed when aggregating relevant effect sizes and outline our general analytic strategy.

Inclusion Criteria

In order to be considered a potential demonstration of sobering (and thus be included in this review), a study had to meet two criteria. First, it had to measure an expectation or prediction regarding a self-relevant outcome. This outcome had to be personal (i.e., had to involve the self directly), so studies on predicting societal events (e.g., elections) or other people's behavior were excluded. We also excluded studies of planning behavior, in which the outcome is a self-generated behavior (or lack thereof) rather than external feedback and thus is not fixed in time. Note that studies *did not* have to provide a comparison between predictions and reality (or a realistic baseline) to be included. Our goal was to examine shifts in predictions over time, not the changing strength of an optimistic bias.

Second, the study had to measure these expectations at two or more points in time either before the performance or between the performance and obtainment of feedback. As discussed earlier, in some studies, participants provided all estimates prior to both performance and feedback, while in others, participants provided estimates following performance but prior to feedback. We examine whether this is a substantive methodological point in the current analysis; however, both approaches have the potential to capture changes in predictions over time (i.e., sobering) and are thus relevant. For the three studies that utilized more than two assessments, only the final two assessments were considered. Finally, studies on task completion times (e.g., Buehler et al., 1994) were excluded given that predictions of completion times do not retain the same meaning across time and the performance and outcome are confounded.

Literature Search

In order to locate relevant research reports, we first consulted our own article databases, reports cited in those articles, and all studies referenced in the most recent review of the relevant literature (Carroll et al., 2006). We also specifically examined all articles that cited Nisan (1972), Gilovich and colleagues (1993), and Shepperd and colleagues (1996), three now classic citations within this literature. Next, we performed a targeted search of abstracts in the PsycINFO database using various combinations of the terms *expectations, predictions*, *bracing, wishful thinking, optimism,* and *pessimism* on one hand and the terms *time, temporal, proximity, feedback, delayed,* and *immediate* on the other. This yielded a total of 26 usable research reports involving 67 separate samples. Finally, we directed an e-mail request for published or unpublished studies that met the above criteria to listservs for the *Society for Personality and Social Psychology*, *Society for Judgment and Decision Making*, and *European Association for Experimental Social Psychology*. These request results in two additional reports (four samples), for a grand total of 28 reports (71 samples). Seven reports (10%) were unpublished.

Coding of Relevant Data

Extracting effect sizes. As mentioned earlier, we only considered studies that measured expectations during at least two time points prior to performance feedback. Our main interest was in the magnitude of the shift in predictions from the earlier to the later time point. Given that researchers have used a variety of designs to study temporal changes in predictions, it was critical that we employ an effect size metric that has an equivalent interpretation regardless of the design that yielded it. To this end, we employed Cohen's *d*, a standardized mean difference between expectations at the earlier (more distant) and later (more proximal) time point utilizing the pooled standard deviation (see Equation 1; Cohen, 1988). Note that Cohen's *d* retains the same meaning regardless of the design (Borenstein, Hedges, Higgins, & Rothstein, 2009), even though inferential tests for repeated measures rely on difference scores and resultantly employ a different error term (Snedecor & Cochran, 1980). This is an innovative feature of the current analyses as it allowed us to examine effects from between and within-subject designs simultaneously. When multiple measures were employed within a sample, we averaged their effect sizes into a single estimate in order to preserve statistical independence between the estimates.

$$d = \frac{M_{distant} - M_{proximal}}{\sqrt{\frac{(n_{distant}-1)SD^2_{distant} + (n_{proximal}-1)SD^2_{proximal}}{n_{distant} + n_{proximal^{-2}}}}} \quad (1)$$

As is typical in meta-analyses, in multiple cases, we had to correct for missing data. For *between-subjects* designs the most common omission was lack of cell sizes, so in such cases we assumed individuals in the sample to be equally distributed across conditions based on the overall sample size (83% of between-subjects cases). In cases for which standard deviations were not available (10% overall), we computed the effect size by estimating the pooled standard deviation from inferential statistics or *p* values per procedures described in Rosenthal (1991). Note that these transformations were used only in case of estimates derived from between-subjects designs, given that in within-subject designs these transformations are inappropriate because auto-correlation of responses influences the variances (see Dunlap, Cortina, Vaslow,

& Burke, 1996). Thus, in relevant cases utilizing within-subject designs, we estimated standard deviations by imputing an average of standard deviations found for similar studies in that research report (7% overall). For cases without comparable studies, we estimated the standard deviation by dividing the theoretical range of the response scale by six (10% overall).

General study characteristics. Note that meta-analyses include multiple research *reports* (e.g., published journal articles), which may contain one or more *studies* (i.e., groups of individuals recruited together), each of which may yield one or more *data points* (in cases where estimates are derived separately for each condition in between-subjects studies). For each research *report* we recorded the author(s), outlet, year of publication, and the means by which the report was located. For each *sample* (one or more of which can be contained in a single research report) we recorded (a) incentive for participation (*money/extra credit/course requirement/volunteer/other*), (b) sample population (*student/other*), (c) country of participation, (d) age of sample, (e) gender composition, (f) setting (*field/laboratory*), (g) design (*between-subjects/within-subject*), and (h) the general context of the study. In the case of between-subjects designs, we derived a separate estimate for each condition, given that many studies manipulated factors of relevance to our analysis. Thus, some samples yielded more than one data point while preserving the statistical independence of the effect sizes. Basic information about individual study samples can be found in Table 3, while information about participants constituting the meta-analytic database is presented in Table 4.

Prediction and event attributes. In this section, we describe moderator information that was coded for each sample in the analysis. We first describe how we coded features of the prediction estimates themselves, including measurement type and timing. We then describe how we coded two features of the performance domains (importance and outcome familiarity).

The first critical feature of predictions was their overall temporal relation to performance, namely, whether the estimates were collected prior to the performance in question (*preparation phase*) or following the performance but preceding feedback (*waiting phase*). Furthermore, we recorded the objective temporal distance between each estimate and *feedback* in the most specific units available for a given study. For each distant prediction we recorded the number of days, hours, and minutes remaining until the feedback, indicating *never* in cases where feedback was not available at all. Similarly, for each proximal prediction we recorded the number of days, hours, and minutes remaining until the feedback, indicating *immediate* when feedback was immediately forthcoming.

In numerous cases (59%) exact information about time remaining to feedback was not available. When the estimate occurred within the study session that included provision of feedback (but was not immediately prior to feedback) we imputed 30 min. When vague descriptors suggested only a small amount of time (e.g., "just a few minutes"), we imputed 5 min. For studies that involved variability in feedback proximity across participants, we imputed the average temporal distance. We then converted all distances to minutes, capping the latency at 6 months (259,200 min) to reduce skew. When the distant time point was *never* we treated the latency as the most extreme distance of 6 months, and when the proximal time point was *immediate* we treated the latency as 1 min. These procedures allowed us to calculate an objective difference in time between the distant and proximal prediction points. In the interest of thoroughness, we also compared average effect sizes of studies that indicated a fixed distant time point and studies in which participants believed they would never receive feedback at the distant time point.

It is unlikely that such objective metrics adequately capture the subjective meaning of temporal differences across prediction points. To supplement this objective information, two coders (one an author and the other a graduate student in psychology blind to hypotheses) judged how far away the feedback would "feel" to participants in a given study by selecting a number between 1 (*near*) to 7 (*far*) on a Likert-type scale. There was substantial agreement regarding subjective distance for both distant ($r = .79$) and proximal ($r = .68$) estimates, so a mean subjective distance score was created by averaging the two judges' ratings. As was the case with objective temporal information, subtracting these estimates yielded an estimate of prediction time lag involved in each study. The objective and subjective indicators of the overall time lag correlated at $r = .61$.

Next, we recorded the type of prediction as either *degree* of performance success (a point prediction, relative ranking, or an expression of confidence) or *likelihood* of performance success (a likelihood or a probability judgment of success or failure). Information regarding prediction timing, distance from feedback, and form of estimate for each individual study is presented in Table 3.

Other critical study characteristics that were of theoretical interest were outcome importance and outcome familiarity. *Outcome importance* was defined as the relevance or consequentiality of the performance for the typical participant in the study and was coded on a 1 (*low*) to 7 (*high*) Likert-type scale. *Familiarity* with the performance domain was defined as the amount of experience

Table 4
Characteristics of Research Participants in the Meta-Analytic Database

Sample attribute	Relative frequency of cases	Range
Sample population		
Students (%)	100	
Other (%)	0	
Country of participation		
USA (%)	82	
Israel (%)	7	
Netherlands (%)	6	
France (%)	4	
Canada (%)	1	
Setting		
Field (%)	32	
Laboratory (%)	68	
Incentive for participation		
Course requirement (%)	54	
Extra course credit (%)	28	
Volunteer (%)	15	
Money (%)	3	
Age of sample (in years, 10% reporting)	19.6	19.1 to 20.3
Gender composition (% female, 51% reporting)	71.9	42 to 100

Note. The relative frequency for each category is only calculated within that category.

the typical participant in the study had with making predictions in the domain, coded on a 1 (*low*) to 7 (*high*) Likert-type scale. As with subjective temporal distance, two coders agreed substantially in their ratings of each domain (*r*s of .66 and .72, respectively), and thus we created mean importance and familiarity scores by averaging the two judges' ratings.

Aggregation of Effect Sizes

When aggregating effect sizes, a typical decision involves choosing between a fixed-effect and a random-effects model. A fixed-effect model assumes a singular population effect underlying all the studies being considered and thus recognizes sampling error (of participants into studies) as the only source of variation in sample estimates of the population effect size (Borenstein et al., 2009; Hedges & Olkin, 1985). This type of model is most accurately called a *constant-coefficient* model, given that a sampled study effect size is considered a function of a single constant population value (Bonett, 2010). Such models strongly weight study effects by their sample size, and often produce overly narrow confidence intervals around the population mean (Bonett, 2010; Hunter & Schmidt, 1990). In other words, when true effect sizes are *not* identical across the studies these confidence intervals have true coverage probability that is much smaller than the reported level of confidence.

Random-effects models, on the other hand, assume multiple effect sizes in the population and acknowledge that the variation in sample estimates is driven not just by (participant) sampling error but also by substantive differences among the studies reflected in the study population (Hunter & Schmidt, 1990). This type of model is most accurately called a *random-coefficient* model, given that a sampled study effect size is considered a function of a population value that is randomly sampled from a normal distribution of multiple population values (whose mean and variance are estimated in order to compute meta-analytic estimates, Bonett, 2010; Hedges & Vevea, 1998; Hunter & Schmidt, 1990). Consequently, these procedures weight individual effect sizes by sample size much less, resulting in wider confidence intervals around the population mean. However, these confidence intervals for the superpopulation mean and variance can be extremely wide and do not perform properly if the superpopulation is not normally distributed or the sampling from this population is not random (Shuster, 2010).

There has been an extensive debate about the appropriateness and robustness of these inferential approaches (Cooper & Hedges, 1994, 2009; Schulze, 2004), and constant refinements to meta-analytic procedures are being proposed. Rather than reviewing all points of controversy, we briefly describe the impetus behind the inferential approach we employed. In order to produce most accurate confidence intervals, we relied on the Synthesizer 1.0 software package (Krizan, 2010), which uses the unweighted least-square estimator recently proposed by Bonett (2008, 2010). This estimator is based on a *varying-coefficient* model, where a sampled study effect size is considered a function of a population value that does vary, but not necessarily along a normal distribution from which it is randomly sampled (Bonett, 2010; see Judge, Griffiths, Hill, Lütkepohl, & Lee, 1985, for discussion of varying-coefficient models). This approach should be preferred because it does not rely on the unrealistic assumption of fixed-effect (i.e., constant coefficient) models that there is no variation in population effect sizes (Hunter & Schmidt, 2000), while avoiding assumptions about a random selection of studies from a normally distributed (and often ill-defined) superpopulation central to random coefficient models (Hedges & Vevea, 1998; Krizan, 2010; Shuster, 2010). In short, the model we employed used an unweighted average of study effect sizes as an index of the population effect[1] (see Bonett, 2008, 2009, for details and extensive simulations showing the advantage of this method compared to fixed-effect and random-effects approaches). In the next section, we provide a summary of relevant computations.

Overview of computations. As indicated earlier, each data point was a standardized mean difference between predictions at the more distant and more proximal time point (Equation 1). Next, we calculated standard errors for each effect size estimate per procedures outlined by Bonett (2009), assuming equal variances for responses in distant and proximal conditions. For estimates from between-subjects designs, the standard error was computed as follows:

$$SE(d_{between-subjects}) = \sqrt{\frac{d^2\left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right)}{8} + \frac{1}{n_1} + \frac{1}{n_2}} \tag{2}$$

For estimates from within-subject designs, we first estimated the auto-correlation between responses given at two time points by using Equation 3 below (Bonett, 2009, p. 227; the *t* value and degrees of freedom are from a repeated-measures *t* test).

$$r = \frac{\left[SD_1^2 + SD_2^2 - (df + 1)(M_1 - M_2)^2/t^2\right]}{(2SD_1SD_2)} \tag{3}$$

We then used this value to calculate the standard error of effect sizes from within-subject designs by using Equation 4 below (Bonett, 2009), also assuming equal variances in the distant and proximal estimates.

$$SE(d_{within-subjects}) = \sqrt{\frac{d^2(1 + r)}{4(n - 1)} + \frac{2(1 - r)}{n}} \tag{4}$$

Note that squaring standard error values in Equations 2 and 4 yields an estimate of effect size *variance*, used in computations of confidence intervals described below. In order to compute confidence intervals around aggregated estimates, we utilized the following formula based on the varying-coefficient model that em-

[1] Common statistical wisdom dictates that large samples exhibit less sampling error and thus should be afforded higher weight when computing aggregated estimates (e.g., Borenstein et al., 2009). However, in a typical meta-analysis, effect sizes originate from subpopulations that may be drastically different in the magnitude of the effect (Hunter & Schmidt, 2000). Furthermore, in a typical meta-analytic database that is relatively small (e.g., less than a 100 samples), the largest samples may yield effects that come from subpopulations from either end of the continuum, biasing overall estimates toward subpopulations that represent either the high- or the low-end population effects (which is one reason why random-effects models confer less weight to sample size information when aggregating estimates and yield larger confidence intervals). Thus, we follow the recommendation that confidence intervals be constructed around *unweighted* averages (Shuster, 2010), which tend to outperform confidence intervals generated based on traditional fixed-effect or even random-effects models (Bonett, 2009, 2010).

ploys an unweighted estimate of the cumulative effect size (see Bonett, 2009; Krizan, 2010, for more detail).

$$\bar{d} \pm Z_{\alpha/2} \sqrt{\frac{\sum_{i=1}^{m} b^2 var(d_i)}{m^2}} \tag{5}$$

Note that this approach allowed pooling of effect sizes from both between- and within-subject designs, as each yielded its own appropriate variance estimate. The coefficient *b* reflects the small sample bias adjustment originally proposed by Hedges (1981), and this was calculated for between- and within-subject designs as presented in Equations 6 and 7, respectively (see Bonett, 2009).

$$b_i = 1 - 3/\{4(n_{1i} + n_{2i}) - 9\} \tag{6}$$

$$b_i = \{(n_i - 2) / (n_i - 1)\}^{1/2} \tag{7}$$

General analytic strategy. In order to estimate the strength of the effect for each group of studies, we computed 95% confidence intervals as presented in Equation 5. We then interpreted the strength and breadth of the effect in question. Addressing the critical questions of the analysis involved comparing effect sizes across various moderating factors. For questions that involved comparing groups of studies along a categorical distinction (e.g., design type), we computed linear contrasts of means using Synthesizer 1.0 (Krizan, 2010), yielding an estimate of the difference in effect sizes and associated 95% confidence interval (Bonett, 2009), expressed in Equation 8.

$$\sum_{i=1}^{m} c_i\hat{\delta}_i \;\pm\; Z_{\alpha/2}\left\{\mathrm{var}\left(\sum_{i=1}^{m} c_i\hat{\delta}_i\right)\right\}^{1/2} \tag{8}$$

When such direct statistical comparisons were inappropriate (i.e., when some studies yielded data points for multiple moderator categories being compared), we simply compared the breadth of confidence intervals for each subgroup, allowing us to make substantive conclusions about the extent to which the effects differed.

For questions that involved continuous moderator variables (e.g., outcome importance), we correlated effect sizes with their corresponding values on the moderating variable (i.e., coder ratings). For an inferential test of this correlation, we used Weighted-Least Square (WLS) regression where we regressed the effect sizes on the variable of interest while using their standard errors as weights (given assumptions of scedasticity do not hold for an inferential test of the correlation). Note that our approach departs from tradition, which often relies on inferential assessments of heterogeneity, for example via the Q statistic (Cochran, 1954). However, this traditional approach often suffers from low power, may yield different results for different effect size metrics, and does not directly speak to the magnitude of the difference (Bonett & Wright, 2007; Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). The current analysis allowed for straightforward interpretation of effect size estimates without reliance on a somewhat arbitrary decision regarding heterogeneity. However, in the interest of thoroughness, we present Q estimates for all groups and group comparisons. Finally, we note that Cohen (1988) recommended that effect sizes of .2, .5, and .8 be tentatively considered as small, medium, and large, although the magnitude of effects should always be considered within their theoretical and practical context (Valentine & Cooper, 2003).

Results

Overview

We begin by offering a general description of the study population and relevant variables, after which we present confidence intervals around population estimates for various subpopulations of studies. We then focus specifically on differences between these estimates as a function of proposed categorical moderators and present direct comparisons where appropriate. For moderators that were coded as continuous variables we present correlations with population effect size estimates. Finally, we discuss the potential impact of publication bias and provide a summary of the main findings.

Characteristics of the Empirical Database

The key characteristics of all studies are presented in Table 3, and features of research participants from the study database are presented in Table 4. The analysis involved a total of 71 samples utilizing 3,451 participants. As is typical for psychological research, a majority of studies were based on predominantly female samples (72% female on average) of university students who received an incentive for their participation. Although overreliance on students generally restricts generalizability, in the present analysis this was often a positive feature as in many cases the research examined students' actual grade predictions in a variety of courses, enhancing the ecological validity of the data. Furthermore, studies employed individuals from five different countries spanning three continents.

Descriptive information about key study features is presented in Table 2. As can be gleaned at the top of the table, the distant prediction occurred 79 days ahead of performance feedback, whereas the proximal prediction occurred 26 days ahead of performance feedback, on average, yielding a temporal lag of around 53 days. As would be expected, these values were substantially smaller for post-performance cases where the proximal predictions occurred generally 1 day before feedback was provided. The raters' judgments of temporal distance were generally consistent with objective metrics of time, and neither varied greatly as a function of the design employed. Pre-performance time lags were objectively shorter than post-performance time lags, although ratings of subjective temporal distance suggest no large differences. In terms of domain importance and familiarity, the prediction domains were generally rated as important and somewhat familiar. Finally, the samples were fairly evenly distributed across key moderating features involving performance timing, experimental design, and measure type.

The Overall Effect

Considered as a whole, the data indicate a robust effect, with the population average estimated at .40 and falling between .36 and .45 (see Table 5). There was a large dispersion in the individual effects ($SD_\delta = .40$) as individual estimates varied from −0.12 to 1.55, although 90% were in the direction of declining predictions. The effect size distribution was slightly leptokurtic (.78, $SE = .56$, $z = .39$, $p = .70$) with a substantial positive skew (1.09, $SE = .12$, $z = 3.81$, $p < .001$). Inspection of traditional heterogeneity indices

Table 5
Results of Quantitative Syntheses for Overall Effect and Main Categorical Moderator Variables

Category	k	N	δ	95% CI	Q	Fail-safe N
Overall	71	3,451	.40	.36, .45	3,339.5*	51,602
Performance timing						
Preparation phase	41	1,769	.49	.42, .56	2,484.2*	16,116
Waiting phase	30	1,682	.35	.28, .43	787.8*	10,012
			Diff = .14 [.02, .26]; $Q_{\text{btw}}(1) = 3.0$, $p = .08$			
Nature of study design						
Between-subjects	34	2,142	.49	.40, .58	2,166.7*	35,008
Within-subjects	37	1,309	.32	.28, .37	294.4*	1,573
			Diff = .17 [.06, .28]; $Q_{\text{btw}}(1) = 19.9$, $p < .001$			
Study context						
Lab	48	2,229	.42	.36, .48	2,413.1*	30,514
Field	23	1,222	.37	.28, .45	717.9*	2,734
			Diff = .05 [−.07, .17]; $Q_{\text{btw}}(1) = 10.6$, $p < .001$			

Note. With regard to the column headings, k = number of independent samples; N = total number of participants; δ = population estimate of the standardized mean difference; 95% CI = 95% confidence interval around mean effect size; Q = Cochran's heterogeneity statistic; Fail-safe N = the required number of studies reporting no difference necessary to render the effect nonsignificant (Rosenthal, 1979). Q_{btw} = inferential statistic of between-groups variance based on a mixed-effects analysis; Diff = linear contrast of difference in subpopulation means with 95% confidence interval appearing in brackets.
* $p < .001$.

confirmed a large amount of variability across effect sizes, with the I^2 statistic (Higgins, Thompson, Deeks, & Altman, 2003) suggesting that 97% of variation may be attributable to meaningful differences between the studies. Note that the large variability of effect sizes and their nonnormal distribution reinforce our choice of the varying-coefficient model as most appropriate, as assumptions for neither fixed-coefficient nor random-coefficient models seem to be met. In short, while there was a substantial tendency for expectations to decline as performance feedback drew nearer, the magnitude of this decline varied greatly across studies.

Three samples (Nisan, 1972; Nussbaum et al., 2006; Sanna, 1999) could be considered outliers as their estimates exceeded 1.40 (the nearest lower estimate was 1.29). Given the positive skew of the distribution and the fact that these samples' characteristics were uniformly hypothesized to increase sobering (i.e., all examined pre-performance estimates in a between-subjects design employing large time interval differences), they were left in the analysis. When excluding these studies the population average dropped only to .34, suggesting that inclusion of these estimates is unlikely to change the substantive conclusions of the analysis.

The Role of Performance Timing and Experimental Design

Critical questions in our analysis involved the impact that performance timing and study design had on the degree of sobering over time. Population estimates for the *preparation* and *waiting phase* shifts in predictions are presented separately toward the top of Table 5. In accordance with our hypotheses, pre-performance shifts were substantially greater, exceeding post-performance shifts by .14 standard deviations. Population estimates for samples employing between- versus within-subjects designs are presented in the middle of Table 5. Again consistent with our hypotheses, the effects were much stronger in between-subjects designs, exceeding those from within-subject designs by .17 standard deviations.[2] Finally, although estimates from lab studies slightly exceeded those from field studies (.05 standard deviations), the confidence interval around the difference included 0, and its width implies a trivial difference.

We also explored the possibility that the difference in effect sizes between within- and between-subjects designs might depend on performance timing and that the difference in effect sizes between pre- and post-performance predictions might depend on design type. Table 6 presents population estimates separately for different combinations of performance timing and design type. Focusing on design differences first, the top of Table 6 reveals that between-subjects designs yielded larger shifts in cases of both pre- and post-performance predictions. However, this difference was more robust for predictions following performance (.23 standard deviations), as the difference for pre-performance predictions was smaller (.10 standard deviations) and had a confidence interval hugging 0.

Turning to performance-timing differences, the bottom of Table 6 reveals that pre-performance shifts were greater for both design types. This difference was larger for within-subjects studies (.24 standard deviations) but covered a relative wide range of values. The difference for between-subjects studies was smaller (.11 standard deviations) but estimated more precisely. Although suffering from lower precision when broken down by design categories, these estimates are fully consistent with the overall results regarding performance timing from Table 5. Note that these general findings cannot easily be explained by appealing to a confound between performance timing and time lag in predictions, as post-

[2] We also examined the relationship between experimental design and effect size controlling for temporal distance at the proximal time point. As apparent in the second row of Table 2, the proximal time point in between-subjects designs was far closer to feedback than in within-subjects designs. However, a multiple regression revealed that the effect of experimental design remained significant ($\beta = .30$, $p = .006$), even after controlling for proximal distance. We also examined the effect of experimental design after controlling for overall temporal lag (i.e., from the distant to the proximal time point) and once again found that the effect remained significant ($\beta = .32$, $p = .004$).

Table 6
Interaction Effects Between Performance Timing and Nature of Study Design

Category	k	N	δ	95% CI	Q
Preparation phase only					
Within-subjects	20	609	.43	.36, .51	161.6*
Between-subjects	21	1,160	.53	.42, .65	2,300.5*
			Diff = .10 [−.06, .26]; $Q_{btw}(1) = 4.34$, $p = .04$		
Waiting phase only					
Within-subjects	17	700	.19	.16, .23	104.7*
Between-subjects	13	982	.42	.29, .55	682.43*
			Diff = .23 [.08, .39]; $Q_{btw}(1) = 14.51$, $p < .001$		
Within-subjects only					
Preparation phase	20	609	.43	.35, .51	161.57*
Waiting phase	17	700	.19	.15, .24	104.7*
			Diff = .24 [.13, .35]; $Q_{btw}(1) = 1.95$, $p = .16$		
Between-subjects only					
Preparation phase	21	1,160	.53	.42, .65	2,300.5*
Waiting phase	13	982	.42	.35, .49	682.4*
			Diff = .11 [−.05, .27]; $Q_{btw}(1) = 0.56$, $p = .45$		

Note. With regard to the column headings, k = number of independent samples; N = total number of participants; δ = population estimate of the standardized mean difference; 95% CI = 95% confidence interval around mean effect size; Q = Cochran's heterogeneity statistic; Q_{btw} = inferential statistic of between-groups variance based on a mixed-effects analysis; Diff = linear contrast of difference in subpopulation means with 95% confidence interval appearing in brackets.
* $p < .001$.

performance time lags were generally *greater* than pre-performance time lags (suggesting a finding opposite from the one observed), and time lags for between-subjects designs were only slightly larger than those for within-subjects designs (see Table 2).

The Role of Measurement Type

In order to test our hypotheses that predictions of degree (e.g., number of items to be solved correctly) are more likely to show declining expectations than are estimates of likelihood (e.g., probability of success), we compared estimates derived from samples employing each type of measure (see Table 7). Because many studies included more than one type of measure, we do not present direct contrasts as the estimates come from overlapping study populations. As expected, predictions of degree showed shifts .14 standard deviations larger than predictions of likelihood, a difference supported by nonoverlapping confidence intervals for effects on each type of measure.

Table 7
Results of Quantitative Syntheses for Effects Within Measurement Types

Measure type	k	N	δ	95% CI	Q	Fail-safe N
Degree	68	3,857	.47	.43, 52	2,792.0*	62,600
Likelihood	19	592	.32	.23, 41	595.9*	2,123

Note. With regard to the column headings, k = number of independent samples; N = total number of participants; δ = population estimate of the standardized mean difference; 95% CI = 95% confidence interval around mean effect size; Q = Cochran's heterogeneity statistic; Fail-safe N = the required number of studies reporting no difference necessary to render the effect nonsignificant (Rosenthal, 1979).
* $p < .001$.

The Role of Prediction Lag, Timing, and Outcome Characteristics

In this section, we present findings regarding hypothesized links between the sobering and prediction timing, outcome importance, and outcome familiarity. All correlations and associated inferential tests are presented in Table 8, both overall and separately for pre- and post-performance predictions. We begin by considering links between sobering and temporal lag.

Regarding temporal lag in performance predictions, we hypothesized that larger time intervals in predictions (i.e., between the distant and proximal performance estimates) would allow for larger declines in expectations. Although weak, there was a trend for the objective temporal difference (in minutes) to predict larger effects (top of Table 8; $r = .09$, $p < .001$). Similarly, subjective temporal difference (based on raters' evaluations) was related to larger effects ($r = .26$, $p = .06$). When examined separately for pre- and post-performance predictions the form of the relationship does not change, although it is stronger post-performance. Two features of the current data may help explain this difference between pre- and post-performance designs. First, temporal lags were objectively greater for post-performance predictions, although this was not reflected in raters' codes (see Table 2). Second, given that the temporal proximity of the final (proximal) estimate to performance likely intensifies declines in expectations (see below), it should not be surprising that post-performance predictions (for which the final estimate is by definition closer to feedback than it would be prior to performance) show larger shifts. Regardless, the data clearly indicate more sobering over longer time intervals.

In the interest of thoroughness, we also compared studies that included a fixed distant time point with studies in which participants at the distant time point believed they would *never* receive feedback (i.e., infinite temporal distance). Unsurprisingly, studies

Table 8
Results of Quantitative Syntheses for Continuous Moderator Variables

Moderator variable	r	p
Overall (k = 71)		
Prediction timing		
Difference, in minutes	.09	.00
Difference, subjective	.26	.06
Proximal distance, in minutes	−.26	.00
Proximal distance, subjective	−.16	.70
Outcome importance	−.29	.01
Outcome familiarity	.10	.23
Preparation phase predictions (k = 41)		
Prediction timing		
Difference, in minutes	.12	.00
Difference, subjective	.23	.14
Proximal distance, in minutes	−.40	.00
Proximal distance, subjective	−.20	.21
Outcome importance	−.37	.01
Outcome familiarity	.09	.00
Waiting phase predictions (k = 30)		
Prediction timing		
Difference, in minutes	.31	.00
Difference, subjective	.51	.00
Proximal distance, in minutes	−.25	.49
Proximal distance, subjective	−.25	.65
Outcome importance	−.15	.37
Outcome familiarity	−.15	.01

Note. With regard to the column headings, r = standardized correlation coefficient relating the moderator variable with effect sizes; p = p value for the significance of this variable as a predictor of effect sizes in a weighted-least square regression analysis with standard errors of effect sizes as weight.

with an infinite distant time point (k = 13) showed greater shifts in predictions (d = 0.37) than did studies with a fixed, noninfinite distant time point (d = 0.028). The confidence interval around the difference suggests a nontrivial difference in magnitude (.05, .12), providing further support for the importance of prediction timing.

We also expected that the proximity of the final estimate to feedback would predict sobering, regardless of the lag between the predictions. As seen in Table 8, less objective distance to performance feedback at the proximal estimate was associated with larger effects ($r = -.26$, $p < .001$). The subjective distance of the proximal estimate showed the same pattern ($r = -.16$, $p = .70$), but the link was weaker and did not approach significance. A similar pattern was observed even after controlling for time lag differences (partial correlations of −.25 and .05, *p*s .04 and .71, respectively). These trends persisted when examined separately for pre- and post-performance predictions (although not significant in the latter case), confirming that making the final prediction close to performance intensifies declines in expectations.

Regarding outcome importance, we had reasons to hypothesize that people could report greater or smaller declines in expectations for important outcomes. As seen in Table 8, there was a substantial tendency for studies that employed more important outcomes to yield *smaller* effects ($r = -.29$, $p = .01$). This relationship held even after taking performance timing and design type into account ($r_p = -.26$, $p = .03$). This tendency was especially apparent for pre-performance predictions ($r = -.37$, $p = .01$).

Regarding outcome familiarity, we also had competing hypotheses, and in fact the data indicate no overall effect of familiarity ($r = .10$, $p = .23$). However, as shown in Table 8, separating studies in which participants made predictions pre-performance versus post-performance paints a different picture. For pre-performance predictions, there was a *positive* relationship between outcome familiarity and effect size, such that people reported greater declines in expectations for more familiar outcomes ($r = .09$, $p < .001$). For post-performance predictions, there was a *negative* relationship between familiarity and effect size, such that people reported greater declines for *less* important outcomes ($r = -.15$, $p = .01$).

Publication Bias

A critical consideration in every quantitative synthesis (or any systematic review, for that matter) is the extent to which the study database on which population estimates are based may itself be biased by publication and reporting practices. No single statistical solution to this problem exists, and assessing the impact of these practices depends on the mechanisms assumed to drive them in the first place. There is some consensus that studies that are small, that are statistically nonsignificant, and that yield small effect sizes are *less* likely to be published or reported within publications (Begg & Berlin, 1988). Based on one or more of these considerations, a variety of approaches have been proposed as means for estimating the presence and impact of publication bias, at the center of which is inspection of the funnel plot (Sutton, 2009).

The plot representing study effect sizes as a function of sample size is shown in Figure 2, and it does approximate the anticipated "funnel" shape. The typical consideration in inspecting funnel plots is the extent to which there are asymmetries toward the bottom of the plot, indicating that smaller studies, particularly those with null or controversial effects, may have been suppressed from the literature. Visual inspection of the plot does reveal such an asymmetry, with the left tail of the distribution being truncated. This suggests that studies that yielded negative effects (i.e., tem-

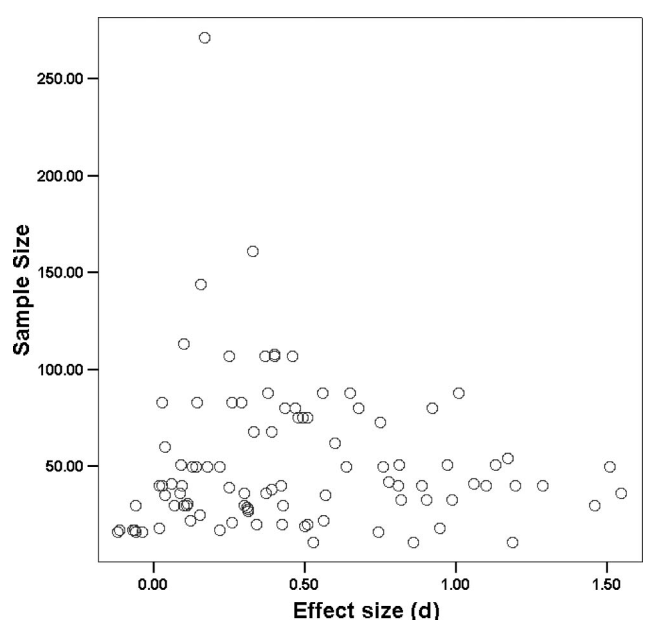

Figure 2. Funnel plot of study effect sizes as a function of sample size.

poral *increases* in expectations) may be missing. One way to quantify this asymmetry is to compute a rank-order correlation between sample size and effect size (Begg & Mazumdar, 1994). This analysis revealed a small correlation that failed to reach conventional standards of significance (Kendall's τ = .07, p = .19). In short, there is some evidence for bias, albeit relatively small.

How best to quantify the potential impact of this bias on the current findings? One straightforward approach is to restrict the analysis to larger studies, assuming they would have been published regardless of their findings (see Sutton, 2009). Aggregating only studies that included over 70 participants (k = 13, see top of Figure 2) yields an overall effect estimate of .38, which is virtually identical to the estimate based on all studies (.40). A related approach is to *trim and fill* the effect sizes, namely, to conduct an iterative procedure in which the most extreme small studies with positive effects are removed ("trimmed") until symmetry in the plot is achieved and are then reintroduced ("filled") twice as exemplifying opposite effects in order to correct the variance (Duval & Tweedie, 2000). Executing this analysis on the current data set (assuming random-effects) yielded an estimate (.26) only .06 standard deviations smaller than the original estimate (.32, based on the random-effects model). Finally, Rosenthal's (1979) fail-safe N statistics (see Tables 5 and 7) for individual population estimates suggest that thousands of studies with null results would be necessary in order render the effects trivial. In sum, we found some evidence that studies yielding smaller effects may be missing, but this suppression effect is likely to be small and does not change the substantive conclusions from the analysis.

Summary of Meta-Analytic Results

In the above analyses, we examined the magnitude of temporal declines in expectations as a function of various factors hypothesized to moderate sobering over time. Consistent with our hypothesis that more factors are likely to contribute to sobering during the preparation phase prior to performance than during the waiting phase following performance, effects were substantially greater for predictions prior to the performance. Furthermore, in line with our hypothesis that consistency-oriented accountability pressures inherent in repeated measures designs are likely to weaken temporal declines in expectations, effects were substantially greater for between- rather than within-subject designs, particularly for post-performance estimates.

In terms of temporal features of predictions, greater time lags between prediction points were associated with larger declines in expectations, as would be expected given more opportunity for various factors to exert a larger cumulative effect. As we anticipated, people showed particularly precipitous declines in expectations when the proximal prediction was made closer to performance feedback (regardless of the time interval between predictions), supporting our notion that greater declines are likely experienced immediately prior to receiving feedback due to rising accountability pressures and affect management concerns. We also tested whether similar processes are more likely to operate when performance outcomes are important or whether desirability biases may shield predictions from precipitous declines in expectations. Our findings indicate that people sobered more over time when outcomes were *less* important, particularly in advance of performance. The relationship between outcome familiarity and sobering differed depending on performance timing: People sobered more for familiar outcomes when making predictions before a performance, but sobered *less* for familiar outcomes when making predictions after a performance. In addition, predictions of degree were more sensitive to temporal declines than were predictions of likelihood. Furthermore, the effects did not vary as a function of experimental setting. Finally, the results were generally robust, and publication bias seemed to play only a minor role in shaping the effect size estimates.

Discussion

We began our inquiry with the aim to establish the strength and robustness of the temporal declines in expectations frequently observed in examinations of performance predictions. Furthermore, we explicated a theoretical proposition that changes in control, construal, accountability pressures, and affect management concerns prompt these declines in expectations (i.e., sobering). We further proposed that these four causes of sobering would manifest in a set of testable moderating variables that either enhance or diminish the tendency to sober up over time. How did our hypotheses fare? The first noteworthy conclusion to arise from our review is that the effect is strong and quite robust. Nearly all of the samples in our review demonstrated a shift in predictions in the expected direction (i.e., a downward shift as feedback drew near), and many of the effects were quite large, with 20% of standardized mean differences exceeding .70. The strength and uniformity of the pattern of sobering over time confirms that expectations are malleable and responsive to subtle changes in perceptions of a performance and performance outcomes, a finding that is particularly striking in light of the presumed ubiquity of optimistic outlooks (S. E. Taylor & Brown, 1988; Peterson, 2000) and the affective and behavioral consequences of expectations (Krizan et al., 2010; Oettingen & Gollwitzer, 2009; Sweeny et al., 2011; Sweeny & Shepperd, 2010).

Although the overall effect of sobering was quite robust, our review revealed key moderators of the effect that were broadly consistent with our hypotheses. Perhaps most importantly, studies that measured performance predictions during the preparation phase *prior* to the performance reported stronger effects than did studies that measured predictions during the waiting phase *after* the performance but prior to feedback. This finding supports the hypothesis that people face more cues to sober up as they prepare for a performance (declines in control over outcomes, shifts in construal of the performance, increasing accountability pressure, and intensifying concerns over affect management) than while they await feedback, when only accountability pressures and affect management concerns continue to influence expectations. A related interpretation of this finding is that perhaps experience with the performance restricts the previously limitless range of possible outcomes (e.g., by revealing weaknesses in one's memory for the test-relevant material), such that predictions during the waiting phase have less "room to move" than predictions during the preparation phase. This explanation is compatible with our hypothesis that people show weaker declines in expectations during the waiting phase because aspects of the situation (namely, control and construal) become fixed once the performance is over.

This moderating effect of performance timing points to a second noteworthy conclusion of our review, which is that all sobering is not created equal. Until now, the literature on shifts in expectations has lumped together studies that target preparation and waiting phase predictions with no comment regarding their equivalence (or lack thereof; e.g., Carroll et al., 2006). The difference in degree of sobering between these two types of studies, as well as the more nuanced finding that the influence of performance timing interacts with other features of forecasts to predict the magnitude of the effect, strongly suggests that the processes driving downward shifts in expectations as people prepare and as they wait are *not* equivalent.

The second key finding was that the effect was generally stronger with greater temporal lags between distant and proximal predictions. In other words, people lowered their expectations more when more time had passed since their initial prediction. The relationship between prediction timing and sobering was consistent in direction across both subjective and objective measures of timing and across pre- and post-performance predictions, although the relationship was strongest for subjective time passage and for studies that measured post-performance predictions. Furthermore, sobering was particularly strong when people believed at the time of their initial prediction that they would never receive feedback and thus experienced an infinite shift in temporal distance from the distant to the proximal time point. These general findings were not only consistent with our hypotheses but perhaps not particularly surprising; after all, the basis for our review was an examination of declines in expectations over time, so one should expect that a greater passage of time would yield greater declines in expectations.

More novel was the finding that studies with proximal predictions occurring closer to feedback showed stronger effects (particularly for post-performance predictions), suggesting that declines in expectations are most precipitous in the final moments before feedback. In fact, this conclusion also sheds light on the previous finding that the relationship between length of temporal lag and sobering was particularly strong for post-performance studies, in that only post-performance studies captured the particularly steep decline in expectations at the moment of truth. This finding highlights the importance of affect management concerns, as anxiety (and likely little else) sharply increases immediately prior to feedback (Shepperd et al., 1996), and thus suggests that the final moments prior to feedback are likely the most painful. Of course, accountability pressure also continue to increase once the performance is complete, but we suspect that the final panicked moments prior to feedback likely leave little time for consideration of cognitively complex concerns over accountability (Easterbrook, 1959; Paulhus & Lim, 1994).

Third, the findings supported our hypothesis that sobering examined via within- and between-subjects designs involves different psychological processes. Although it may be easier to detect statistically significant changes in predictions in within-subjects designs due to their greater power, our examination of the magnitude of these changes reveals that between-subjects studies show stronger effects. This finding confirms our hypothesis that within-subjects designs, in which participants make multiple performance predictions, introduce a type of accountability pressure that leads people to remain at least somewhat consistent in their predictions across measurement points. People who report their expectations on multiple occasions experience a type of behavioral commitment when they record their early outcome predictions, and this early commitment can then strengthen confidence in their initial prediction and reduce the likelihood that they will make a different prediction later (e.g., Davis, 1979). This finding implies that one strategy to maintain optimism in anticipation of feedback is to make an early and optimistic prediction about one's likely outcomes. We also examined the effect of study design separately for pre- and post-performance predictions and found a pattern that confirms our hypotheses for both performance timing and study design: The strongest shifts in expectations occurred in studies that measured predictions prior to the performance and between subjects, and the weakest shifts occur in studies that measured predictions after the performance and within subjects.

Fourth, we examined the possibility that the nature of the prediction measure might be related to the strength of the effect. Specifically, we contrasted predictions of degree (point predictions, relative rankings, and Likert-type measures of success or confidence) with likelihood predictions (probability judgments of success or failure) and found that, as hypothesized, predictions of degree were more sensitive to sobering than were likelihood predictions. We suspected that because predictions of degree are more falsifiable than likelihood predictions (Windschitl et al., 2010), people would feel greater pressure to shift predictions of degree in response to rising accountability and affect management concerns. That is, when people must commit to a specific performance prediction, they face a far greater risk of feedback proving them definitively inaccurate and thus of facing embarrassment and disappointment if they overestimated their outcomes. Our findings suggest that these concerns serve as cues to prompt downward revisions in degree predictions as feedback draws near. In addition, this finding points to another strategy to maintain optimism as feedback draws near: Avoid overly specific outcome predictions in favor of statements of likelihood.

Fifth, we compared effects between field studies and lab studies and were encouraged to discover little difference in sobering between studies conducted in the field (e.g., using real course exam grades) and in the lab. It is always wise to confirm the findings of lab-based studies in the "real world," but it is also useful to know that lab studies examining sobering successfully captured the same intensity of temporal shifts toward pessimism as did real-world studies. This suggests that laboratory studies induce a similar psychological experience of awaiting performance feedback as do naturalistic settings.

Sixth, we examined the effect of outcome familiarity on sobering and uncovered a complex relationship. Although effects were generally equivalent in studies that used familiar and unfamiliar performance domains when taking the set of studies as a whole, a separate examination of studies that assessed predictions before versus after a performance revealed that the overall null relationship obscured a more nuanced relationship between declines in expectations and outcome familiarity. In fact, when people make predictions during the preparation phase, they show greater declines over time for familiar outcomes. Though speculative, one explanation for this finding is that people who are more familiar with a domain are more attuned to the cues in their environment that indicate declining control. For example, a student taking her first college course might not realize that missing class and falling behind in the readings should prompt a downward revisions in her expectations for her midterm performance, whereas an experienced college student would be fully aware of the implications of these choices for the upcoming midterm.

In contrast, when people make predictions following a performance during the waiting phase, they show greater declines over time for *unfamiliar* outcomes. Although this finding seems contradictory to the finding for pre-performance predictions, our theoretical distinction between reasons for sobering before versus after a performance can help explain the seemingly contradictory findings for outcome familiarity. As just discussed, the most likely explanation for greater pre-performance declines for familiar outcomes is that familiar outcomes have familiar cues to indicate declining control, which is a reason for sobering that disappears once the performance is over. After a performance, the primary reason to lower expectations is in response to affect management concerns, which might exert a stronger effect on people who are *unfamiliar* with the performance domain and thus less confident in their outcome predictions. To return to the example of the two students, the more experienced student may be confident about the nature of her performance (whether good or bad) after seeing the exam, based on the many college exams she has taken before, and thus her expectations will remain relatively stable once the exam is over. The inexperienced student may have little sense of her performance even after the exam is over, and thus she may be swayed in her expectations by rising anxiety and a desire to avoid disappointment that would result from overly optimistic expectations. That being said, the observed trends regarding familiarity were relatively weak, and our tentative interpretations await more direct empirical testing.

Finally, we examined the effect of outcome importance and found that people sober up *less* when the performance outcome is more personally important to them, although this finding was more robust for pre-performance predictions. To be clear, the finding is not that people make more optimistic predictions for important outcomes (as in the desirability bias; Krizan & Windschitl, 2007) but rather that the decline in expectations over time is *shallower* when outcomes are more important. We were unsure at the outset whether outcome importance would strengthen or weaken the tendency to sober over time, so any interpretation of this finding must be somewhat tentative. However, this finding might suggest that people "hang on" to high expectations for important outcomes to take advantage of optimism's benefits for motivation and self-efficacy (Armor & Taylor, 1998; Bandura, 1982; Oettingen & Gollwitzer, 2009), an explanation that gains credence from the weaker relationship for post-performance predictions (i.e., when control, and thus benefits of motivation and self-efficacy, is largely gone). Furthermore, individuals may be invested in maintaining an emotionally satisfying outlook of the future (Armor et al., 2008; Massey, Simmons, & Armor, 2011) or may be seeking and evaluating information with an eye toward confirming their hopeful expectations (Krizan & Windschitl, 2007). A key direction for future research should be to take a closer look at temporal trends in expectations for outcomes that vary in their relevance to forecasters' identities or to practical consequences for forecasters.

Implications

We have already noted two key implications of our review. First, tendency to sober over time is strong, robust, and common, suggesting that optimism may not be as ubiquitous as some research suggests (e.g., Armor & Taylor, 2002; Buehler et al., 1997; Weinstein, 1980). In fact, although people may embrace an optimistic outlook when feedback is distant or not expected at all, people readily shift away from these high expectations when they face inevitable and impending feedback. Many writers have argued that optimistic outlooks are adaptive, in that hopeful expectations may feel good (Armor et al., 2008), may project a desirable self-image to others (Pezzo et al., 2006; Tyler & Rosier, 2009), and are likely to promote a focus on planning and investment of control over important outcomes (Bandura, 1982; Oettingen & Gollwitzer, 2009). On the whole, these "positive illusions" are thought to be a hallmark of psychological adjustment and interpersonal success (S. E. Taylor & Brown, 1988). However, this image of people as unabashed optimists dissipates when we consider individuals' outlooks in the moments prior to performance feedback. As our review demonstrates, expectations typically turn toward gloom as the "moment of truth" draws near, especially in the last instants prior to feedback.

The readiness with which people turn their backs on high expectations in the face of feedback suggests that adaptive qualities of optimistic outlooks are more circumscribed than previously thought. Although it may be adaptive to hold hopeful expectations about one's outcomes, it seems equally adaptive to relinquish these expectations in the moments prior to performance or performance feedback. Prior to the performance itself, a more sober outlook could avoid pitfalls of overconfidence such as withdrawal of attention or effort (D. P. Johnson, 2004; Stone, 1994; Vancouver, Thompson, Tischner, & Putka, 2002). Furthermore, after performance but prior to feedback, a sober outlook may set a flattering standard for evaluating one's performance. In fact, recognizing the benefits of declines in expectations has critical implications for how people prepare themselves for potential bad news and how they respond if bad news comes. Emotional responses to feedback depend in large part on a comparison between expectations and outcomes (Mellers, Schwartz, Ho, & Ritov, 1997; Sweeny & Shepperd, 2010; Zeelenberg, van Dijk, Manstead, & van der Pligt, 2000), and people can (and do) capitalize on this psychological phenomenon by lowering their expectations in the face of feedback (Armor et al., 2008).

In contrast to the emotional benefits of sobering is the recent finding that sobering might undermine people's motivation to improve their behavior in response to negative feedback. One study found that people who received better health feedback than anticipated subsequently reported weaker intentions to improve their health relative to people who received unexpectedly bad health feedback (Sweeny et al., 2011). Thus, it seems that sobering may have both beneficial and harmful consequences, and the findings from our review can serve as an impetus to take a closer look at the consequences of this robust phenomenon. For now, we can confidently conclude that sobering is a relatively common experience across numerous domains, which strongly implies that it serves an adaptive function. We can reconcile our findings with the clear benefits of optimism by noting that the advantages of sobering are limited to situations in which feedback is forthcoming and in the relatively near future (Sweeny et al., 2006).

The second key implication from our review is that all sobering is not created equal. That is, we found a robust tendency for people to lower their expectations over time, but the magnitude of this shift depends on when people make performance predictions, the importance of the performance domain, how many times people predict their performance outcomes, and how the predictions are measured. We proposed a set of theoretical explanations for sobering (changes in control, construal, accountability pressure, and affect management concerns) that served as the basis for our hypotheses, and with few exceptions, our findings confirmed the usefulness of this theoretical approach. Moreover, with the exception of outcome importance, our analysis is the first to examine the role of these theoretically derived moderators of sobering.

Our findings provide suggestive evidence for the unique importance of each of the four theoretical explanations, not simply the group of explanations as a whole. The finding that sobering is weaker following a performance (but prior to feedback) reveals the collective importance of declines in control and shifts in construal, which exert influence only before the performance occurs. Although our review could not provide clear evidence for a unique role of shifts in construal, it does tentatively suggest that declines in control are uniquely influential. We found that people lowered their predictions less for important outcomes particularly prior to performance, and our interpretation of that finding suggests that people may cling to optimism when it can still serve to motivate useful behavior (i.e., when opportunities for control remain). Similarly, the finding that people showed greater declines in expectations for familiar outcomes prior to a performance also points to the key role of declining control (and in this case, attention to cues indicating declining control).

Turning to concerns over affect management and accountability pressures, their collective influence appears most clearly in the strength of effect for post-performance predictions, when they are the sole cues to shift predictions downward. These two concerns also best explain the larger shifts in predictions of likelihood as opposed to degree, which introduce heightened concerns over affect and accountability (here, accountability to accuracy) due to their relative falsifiability. Furthermore, the particular strength of sobering in the moments closest to feedback points to a key role of rising anxiety, and the smaller shifts in predictions in within-subjects designs likely reflect the particular influence of accountability pressure (in this case, pressure to remain consistent with previous predictions).

How might the implications of our review apply to predictions that were outside the scope of our inquiry? Most notably, we restricted our review to studies that examined predictions for personal performances and thus did not include predictions of external outcomes such as political elections, sporting events, or financial markets. Further research could draw direct comparisons between the prediction processes for personal and external events, but our review suggests some reasons why these predictions might follow different paths over time. Although construal of external events likely changes over time (Hunt, Kim, Borgida, & Chaiken, 2010), control, accountability pressures, and affect management concerns may change little or not at all. Control over the outcomes of external events is by definition out of the hands of the predictor, and people may feel less accountable for their predictions of external events and less concern over the affective consequences of inaccurate predictions. Of course, some external events have greater personal consequences than others (for example, if people volunteer for a political campaign or have a strong affiliation with a sports team), and we suspect that as personal consequences intensify so will potential accountability and affect management concerns, ultimately prompting sobering for external events.

Of course, expectancy-relevant information about external events is often intentionally managed by other individuals invested in the outcomes of such events. In the case of political campaigns, the public face of a campaign often promotes optimism among supporters of a given candidate, even when polling data suggests the candidate may be doomed. This strategy plays a key role in motivating voter turnout and maintaining a chance at electoral victory (Fenwick, Wiseman, Becker, & Heiman, 1982). Critically, these efforts may promote the maintenance of an optimistic outlook and overshadow any sobering over time among voters. In the 2008 U.S. presidential election, for instance, supporters of Senator John McCain remained somewhat optimistic about McCain's chances for victory all the way up to Election Day (Gallup Poll, 2009; Krizan et al., 2010). In sum, the tendency to sober over time is likely more fickle in domains where information is controlled by external agents invested in the outcome, such as elections or athletic competitions.

Finally, although our discussion focuses on situational predictors of sobering, it is worth noting that several individual differences may also be relevant to these changes. Perhaps the most obvious example is dispositional optimism, which captures the extent to which people have a generalized positive outlook on their future outcomes (Scheier & Carver, 1985). Although it makes intuitive sense that people who generally expect the best would also make more positive predictions for specific future outcomes, in fact, dispositional optimism is typically unrelated to specific outcome predictions (Radcliffe & Klein, 2002). Furthermore, there currently is no evidence that dispositional optimism predicts declines in expectations over time, and multiple findings from our own laboratories suggest that dispositional optimism is unrelated to sobering. These findings (or lack thereof) suggest that the powerful situational characteristics of awaiting feedback override generally positive or negative future outlooks to prompt temporary downward shifts in specific outcome predictions.

Unlike dispositional optimism, defensive pessimism and strategic optimism are more likely to interplay with situational variables to predict sobering. Defensive pessimism and strategic optimism are two contrasting strategies by which people manage their expectations prior to a performance over which they have control. Specifically, defensive pessimists deliberately set low expectations before engaging in a task or performance, which creates anxiety and prompts them to engage in preparative behaviors, whereas strategic optimists maintain high expectations to avoid distracting anxiety over an upcoming performance (Norem & Cantor, 1986; Spencer & Norem, 1996). Like some cases of sobering, defensive pessimism is a strategic, goal-oriented lowering of expectations, albeit one that is only useful prior to a performance when preparative behaviors can still influence defensive pessimists' outcomes. Although we know of no studies that have examined the relationship between defensive pessimism (or strategic optimism) and downward revisions in predictions over time, we suspect that people who engage in these preparative strategies experience different trajectories in their pre-performance predictions. However, the nature of these trajectories is somewhat difficult to predict. It is possible that defensive pessimists would show particularly precipitous declines in pre-performance expectations as time runs out, but it is also possible that defensive pessimists would experience shallower shifts in expectations due to the declining usefulness of pessimism as the performance draws near. Future studies should directly examine individual differences in the trajectories of predictions over time, but for now the preponderance of evidence points to the strength of the situation in prompting sobering.

Limitations and Challenges

Before concluding, a comment on the limitations of our review as well as the challenges facing this area of inquiry is necessary. A strength of our review is its attention to theoretically derived hypotheses; however, we tested these hypotheses meta-analytically. Given that meta-analysis ultimately is a descriptive technique (Victor, 1995), it is important to confirm these hypothesis through direct experimentation and observation. As indicated throughout our review, however, existing studies provide direct evidence for many of our theoretically derived proposals. Furthermore, it is possible that idiosyncrasies of studies within the literature may have biased our conclusions (e.g., overreliance on predictions of academic performances). However, we would note that the same limitation is true for any individual study or review. The value of the current analysis lies precisely in the fact that we found broad-based support for the impact of theoretically relevant psychological processes and their constraining factors on temporal declines in expectations.

With that said, it remains unclear to what extent sobering in a particular context or study reflects any specific process or processes (i.e., changes in control, construal-level, accountability, or affective concerns). Our review provides some insight into both the overall influence of these psychological processes and the relative importance of each process across differing circumstances, but longitudinal studies that assess changes in these processes over time are necessary to pin down their role more precisely. Fortunately, our findings point to multiple factors that should take center stage in future examination of these temporal changes. As indicated earlier, the extent to which outcome importance affects temporal trends in expectations and their affective and behavioral consequences should be a key avenue for future research, along with studies that target outcome familiarity to examine our post hoc interpretation for the distinct effects with pre- and post-performance predictions. Another unresolved question is how people resolve accountability-related conflicts between maintaining consistency with prior predictions and pursuing consistency with upcoming feedback. Finally, the literature on sobering remains relatively silent with regard to the accuracy of people's changing expectations over time. Some studies confirm a shift from unrealistic optimism toward realism or even unrealistic pessimism (e.g., Shepperd et al., 1996), but the majority of studies in this area do not provide an objective baseline or clear outcome to which predictions can be compared. Pursuit of these inquiries will be critical for developing a deeper understanding of the causes and consequences of sobering over time.

Coda

People tend to be optimistic creatures, looking forward to a long life, imagining it full of pleasures and successes, and savoring the achievements that are yet to come. Numerous writers have noted this to be a useful affair in that it helps people deal with personal setbacks and provides people with resolve to continue pursuing their dreams. When facing the moment of truth, however, people often abandon their rosy outlook. The realization that time has run out, that one's perception was skewed, that others may witness one's incompetence or blindness, or that disappointment may be right around the corner, all conspire to prompt awareness that the future may not be as bright as initially hoped.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Armor, D. A., Massey, B. C., & Sackett, A. M. (2008). Prescribed optimism: Is it right to be wrong about the future? *Psychological Science, 19,* 329–331. doi:10.1111/j.1467-9280.2008.02089.x

*Armor, D. A., & Sackett, A. M. (2006). Accuracy, error, and bias in predictions for real versus hypothetical events. *Journal of Personality and Social Psychology, 91,* 583–600. doi:10.1037/0022-3514.91.4.583

Armor, D. A., & Taylor, S. E. (1998). Situated optimism: Specific outcome expectancies and self-regulation. *Advances in Experimental Social Psychology, 30,* 309–379. doi:10.1016/S0065-2601(08)60386-X

Armor, D. A., & Taylor, S. E. (2002). When predictions fail: The dilemma of unrealistic optimism. In T. Gilovich, D. Griffin,, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 334–347). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511808098.021

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37,* 122–147. doi:10.1037/0003-066X.37.2.122

Baumeister, R. F., & Showers, C. J. (1986). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *European Journal of Social Psychology, 16,* 361–383. doi:10.1002/ejsp.2420160405

Begg, C. B., & Berlin, J. A. (1988). Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society: Series A. Statistics in Society, 151,* 419–463. doi:10.2307/2982993

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50,* 1088–1101. doi:10.2307/2533446

Bonett, D. G. (2008). Meta-analytic interval estimation for bivariate correlations. *Psychological Methods, 13,* 173–181. doi:10.1037/a0012868

Bonett, D. G. (2009). Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychological Methods, 14,* 225–238. doi:10.1037/a0016619

Bonett, D. G. (2010). Varying coefficient meta-analytic methods for alpha reliability. *Psychological Methods, 15,* 368–385. doi:10.1037/a0020142

Bonett, D. G., & Wright, T. A. (2007). Comments and recommendations regarding the hypothesis testing controversy. *Journal of Organizational Behavior, 28,* 647–659. doi:10.1002/job.448

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. New York, NY: Wiley. doi:10.1002/9780470743386

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129–148. doi:10.1037/0003-066X.36.2.129

Buehler, R., Griffin, D., & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychology Bulletin, 23,* 238–247. doi:10.1177/0146167297233003

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. doi:10.1037/0022-3514.67.3.366

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

*Carroll, P. J., Shepperd, J. A., & Arkin, R. M. (2009). Downward self-revision: Erasing possible selves. *Social Cognition, 27,* 550–578. doi:10.1521/soco.2009.27.4.550

Carroll, P., Sweeny, K., & Shepperd, J. A. (2006). Forsaking optimism. *Review of General Psychology, 10,* 56–73. doi:10.1037/1089-2680.10.1.56

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. S. Wyer Jr., & T. K. Srull (Eds.), *Handbook of social cognition: Vol. 1. Basic processes* (2nd ed., pp. 323–417). Hillsdale, NJ: Erlbaum.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10,* 101–129. doi:10.2307/3001666

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

Cooper, H. (1998). *Synthesizing research*. Thousand Oaks, CA: Sage.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York, NY: Russell Sage Foundation.

Cooper, H., & Hedges, L. V. (2009). Research synthesis as a scientific process. In H. Cooper, L. V. Hedges,, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 3–16). New York, NY: Russell Sage Foundation.

Davis, M. H. (1979). Changes in evaluative beliefs as a function of behavioral commitment. *Personality and Social Psychology Bulletin, 5,* 177–181. doi:10.1177/014616727900500210

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177. doi:10.1037/1082-989X.1.2.170

Duval, S., & Tweedie, R. (2000). A nonparametric “trim and fill” method of assessing publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89–98.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66,* 183–201. doi:10.1037/h0047707

Fenwick, I., Wiseman, F., Becker, J. F., & Heiman, J. R. (1982). Classifying undecided voters in pre-election polls. *The Public Opinion Quarterly, 46,* 383–391. doi:10.1086/268735

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin, 117,* 39–66. doi:10.1037/0033-2909.117.1.39

Gallup Poll. (2009). *Election 2008*. Retrieved on March 2, 2009, from http://www.statmodel.com/www.gallup.com/poll/election2008.aspx

*Gilovich, T., Kerr, M., & Medvec, V. H. (1993). Effect of temporal perspective on subjective confidence. *Journal of Personality and Social Psychology, 64,* 552–560. doi:10.1037/0022-3514.64.4.552

Gollwitzer, P. M., & Bayer, U. (1999). Deliberative versus implemental mindsets in the control of action. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 403–422). New York, NY: Guilford Press.

Hedges, L. V. (1981). Distribution theory for Glass’s estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128. doi:10.2307/1164588

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504. doi:10.1037/1082-989X.3.4.486

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal, 327,* 557–560. doi:10.1136/bmj.327.7414.557

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods, 11,* 193–206. doi:10.1037/1082-989X.11.2.193

Hunt, C. V., Kim, A., Borgida, E., & Chaiken, S. (2010). Revisiting the self-interest versus values debate: The role of temporal perspective. *Journal of Experimental Social Psychology, 46,* 1155–1158.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8,* 275–292. doi:10.1111/1468-2389.00156

*Jiga-Boy, G. M. (2008). *Adaptive thinking about the future: Temporal construal, health-related behaviour, and perceived temporal distance* (Unpublished doctoral dissertation). University of Grenoble II, Grenoble France.

Johnson, D. P. (2004). *Overconfidence and war: The havoc and glory of positive illusions*. Cambridge, MA: Cambridge University Press.

Johnson, E., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology, 45,* 20–31. doi:10.1037/0022-3514.45.1.20

Judge, G. G., Griffiths, W. E., Hill, R. C., Ĺutkepohl, H., & Lee, T. C. (1985). *The theory and practice of econometrics* (2nd ed.). New York, NY: Wiley.

*Kettle, K. L., & Häubl, G. (2010). Motivation by anticipation: Expecting rapid feedback enhances performance. *Psychological Science, 21,* 545–547. doi:10.1177/0956797610363541

*Krizan, Z. (2008). *The public side of (over) optimism* (Unpublished doctoral dissertation). University of Iowa, Iowa City.

Krizan, Z. (2010). Synthesizer 1.0: A varying-coefficient meta-analytic tool. *Behavior Research Methods, 42,* 863–870. doi:10.3758/BRM.42.3.863

Krizan, Z., Miller, J. C., & Johar, O. (2010). Wishful thinking in the 2008 U.S. presidential election. *Psychological Science, 21,* 140–146. doi:10.1177/0956797609356421

Krizan, Z., Scherr, K., & Windschitl, P. D. (2011). *Performance predictions as a function of social context*. Unpublished manuscript.

Krizan, Z., & Sweeny, K. (in press). Causes and consequences of expectation trajectories: “High” on optimism in a public referendum. *Psychological Science*.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin, 133,* 95–121. doi:10.1037/0033-2909.133.1.95

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125,* 255–275. doi:10.1037/0033-2909.125.2.255

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. doi:10.1037/0022-3514.75.1.5

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127,* 267–286. doi:10.1037/0033-2909.127.2.267

Massey, C., Simmons, J., & Armor, D. A. (2011). Hope over experience: Desirability and the persistence of optimism. *Psychological Science, 22,* 274–281. doi:10.1177/0956797610396223

McKenna, F. P., & Myers, L. B. (1997). Illusory self-assessments—Can they be reduced? *British Journal of Psychology, 88,* 39–51. doi:10.1111/j.2044-8295.1997.tb02619.x

Mellers, B., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science, 8,* 423–429. doi:10.1111/j.1467-9280.1997.tb00455.x

*Monga, A., & Houston, M. J. (2006). Fading optimism in products: Temporal changes in expectations about performance. *Journal of Marketing Research, 43,* 654–663. doi:10.1509/jmkr.43.4.654

Morling, B., & Evered, S. (2006). Secondary control reviewed and defined. *Psychological Bulletin, 132,* 269–296. doi:10.1037/0033-2909.132.2.269

*Nisan, M. (1972). Dimension of time in relation to choice behavior and achievement orientation. *Journal of Personality and Social Psychology, 21,* 175–182. doi:10.1037/h0032245

Norem, J. K., & Cantor, N. (1986). Defensive pessimism: Harnessing anxiety as motivation. *Journal of Personality and Social Psychology, 51,* 1208–1217. doi:10.1037/0022-3514.51.6.1208

*Nussbaum, S., Liberman, N., & Trope, Y. (2006). Predicting the near and distant future. *Journal of Experimental Psychology: General, 135,* 152–161. doi:10.1037/0096-3445.135.2.152

Oettingen, G., & Gollwitzer, P. M. (2009). Making goal pursuit effective: Expectancy-dependent goal setting and planned goal striving. In J. P. Forgas, R. Baumeister,, & D. M. Tice (Eds.), *The psychology of self-regulation: Cognitive, affective, and motivational processes* (pp. 127–146). Philadelphia, PA: Psychology Press.

Oettingen, G., Pak, H., & Schnetter, K. (2001). Self-regulation of goal-setting: Turning free fantasies about the future into binding goals. *Journal of Personality and Social Psychology, 80,* 736–753. doi:10.1037/0022-3514.80.5.736

Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). Expectancies. In E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 211–238). New York, NY: Guilford Press.

Paulhus, D. L., & Lim, D. T. K. (1994). Arousal and evaluative extremity in social judgments: A dynamic complexity model. *European Journal of Social Psychology, 24,* 89–99. doi:10.1002/ejsp.2420240107

*Pennington, G. L., & Roese, N. J. (2003). Regulatory focus and temporal distance. *Journal of Experimental Social Psychology, 39,* 563–576. doi:10.1016/S0022-1031(03)00058-1

Peterson, C. (2000). The future of optimism. *American Psychologist, 55,* 44–55. doi:10.1037/0003-066X.55.1.44

Pezzo, S. P., Pezzo, M. V., & Stone, E. R. (2006). The social implications of planning: How public predictions bias future plans. *Journal of Experimental Social Psychology, 42,* 221–227. doi:10.1016/j.jesp.2005.03.001

Radcliffe, N. M., & Klein, W. P. M. (2002). Dispositional, unrealistic, and comparative optimism: Differential relations with the knowledge and processing of risk information and beliefs about personal risk. *Personality and Social Psychology Bulletin, 28,* 836–846.

*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. doi:10.1037/h0076300

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641. doi:10.1037/0033-2909.86.3.638

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. New York, NY: Sage.

Rothbaum, F., Weisz, J. R., & Snyder, S. S. (1982). Changing the world and changing the self: A two-process model of perceived control. *Journal of Personality and Social Psychology, 42,* 5–37. doi:10.1037/0022-3514.42.1.5

*Sanna, L. J. (1999). Mental simulations, affect, and subjective confidence: Timing is everything. *Psychological Science, 10,* 339–345. doi:10.1111/1467-9280.00165

*Sanna, L. J., & Schwarz, N. (2004). Integrating temporal biases: The interplay of focal thoughts and accessibility experiences. *Psychological Science, 15,* 474–481. doi:10.1111/j.0956-7976.2004.00704.x

*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. doi:10.1177/0146167298245008

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology, 4,* 219–247. doi:10.1037/0278-6133.4.3.219

Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Lower Saxony, Germany: Hogrefe & Huber.

Severynse, M. (1995). *Webster's II new college dictionary*. New York, NY: Houghton Mifflin.

*Shepperd, J. A. (2010). *Bracing for breast cancer risk feedback*. Unpublished data, University of Florida, Gainesville.

Shepperd, J. A., Findley-Klein, C., Kwavnick, K. D., Walker, D., & Perez, S. (2000). Bracing for loss. *Journal of Personality and Social Psychology, 78,* 620–634. doi:10.1037/0022-3514.78.4.620

*Shepperd, J. A., Grace, J., Cole, L. J., & Klein, C. (2005). Anxiety and outcome predictions. *Personality and Social Psychology Bulletin, 31,* 267–275. doi:10.1177/0146167204271322

Shepperd, J. A., & McNulty, J. K. (2002). The affective consequences of expected and unexpected outcomes. *Psychological Science, 13,* 85–88. doi:10.1111/1467-9280.00416

*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. doi:10.1037/0022-3514.70.4.844

Shuster, J. J. (2010). Empirical vs. natural weighting in random effects meta-analysis. *Statistics in Medicine, 29,* 1259–1265. doi:10.1002/sim.3607

Snedecor, G. W., & Cochran, W. G. (1980). *Statistical methods* (7th ed.). Ames: Iowa State University Press.

Spencer, S. M., & Norem, J. K. (1996). Reflection and distraction: Defensive pessimism, strategic optimism, and performance. *Personality and Social Psychology Bulletin, 22,* 354–365. doi:10.1177/0146167296224003

Stone, D. N. (1994). Overconfidence in initial self-efficacy judgments: Effects on decision processes and performance. *Organizational Behavior and Human Decision Processes, 59,* 452–474. doi:10.1006/obhd.1994.1069

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges,, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis (2nd ed.,* pp. 435–452). New York, NY: Russell Sage Foundation.

*Sweeny, K. (2010a). *The benefits of bracing*. Unpublished data, University of California, Riverside.

*Sweeny, K. (2010b). *Bracing as anticipatory coping*. Unpublished data, University of California, Riverside.

*Sweeny, K. (2010c). *Insufficient bracing*. Unpublished data, University of California, Riverside.

Sweeny, K., Carroll, P. J., & Shepperd, J. A. (2006). Thinking about the future: Is optimism always best? *Current Directions in Psychological Science, 15,* 302–306. doi:10.1111/j.1467-8721.2006.00457.x

*Sweeny, K., & Dillard, A. J. (2010). *The consequences of shifts in expectations for behavioral intentions*. Unpublished data, University of California, Riverside.

Sweeny, K., Dillard, A. J., & Fox, A. (2011). *The effect of expectation disconfirmation on behavioral intentions: Implications for unrealistic optimism and bracing for bad news*. Unpublished manuscript.

*Sweeny, K., & Shepperd, J. A. (2007). Do people brace sensibly? Risk judgments and risk prevalence. *Personality and Social Psychology Bulletin, 33,* 1064–1075. doi:10.1177/0146167207301024

Sweeny, K., & Shepperd, J. A. (2010). The costs of optimism and the benefits of pessimism. *Emotion, 10,* 750–753. doi:10.1037/a0019016

*Sweeny, K., Shepperd, J. A., & Carroll, P. J. (2009). Expectations for others' outcomes. *Personality and Social Psychology Bulletin, 35,* 160–171. doi:10.1177/0146167208327050

*Taylor, K. M., & Shepperd, J. A. (1998). Bracing for the worst: Severity, testing, and feedback timing as moderators of the optimistic bias. *Per-*

sonality and Social Psychology Bulletin, 24, 915–926. doi:10.1177/0146167298249001

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193–210. doi:10.1037/0033-2909.103.2.193

*Terry, M., & Shepperd, J. A. (2004). *Changing standards: Adjusting the definition of an acceptable outcome* (Unpublished master's thesis). University of Florida, Gainesville.

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. doi:10.1037/0022-3514.52.4.700

Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology, 57,* 632–640. doi:10.1037/0022-3514.57.4.632

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110,* 403–421. doi:10.1037/0033-295X.110.3.403

Tyler, J. M., & Rosier, J. G. (2009). Examining self-presentation as a motivational explanation for comparative optimism. *Journal of Personality and Social Psychology, 97,* 716–727. doi:10.1037/a0016231

Valentine, J. C., & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse.

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology, 57,* 660–671. doi:10.1037/0022-3514.57.4.660

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87,* 506–516. doi:10.1037/0021-9010.87.3.506

van Dijk, W. W., & van der Pligt, J. (1997). The impact of probability and magnitude of outcome on disappointment and elation. *Organizational Behavior and Human Decision Processes, 69,* 277–284. doi:10.1006/obhd.1997.2688

*van Dijk, W. W., Zeelenberg, M., & van der Pligt, J. (2003). Blessed are those who expect nothing: Lowering expectations as a way of avoiding disappointment. *Journal of Economic Psychology, 24,* 505–516. doi:10.1016/S0167-4870(02)00211-8

Victor, N. (1995). "The challenges of meta-analysis": Discussion, indications, and contra-indications for meta-analysis. *Journal of Clinical Epidemiology, 48,* 5–8. doi:10.1016/0895-4356(94)00107-2

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39,* 806–820. doi:10.1037/0022-3514.39.5.806

*Wicker, F. W., Turner, J. E., Reed, J. H., McCann, E. J., & Do, S. L. (2004). Motivation when optimism declines: Data on temporal dynamics. *The Journal of Psychology: Interdisciplinary and Applied, 138,* 421–432. doi:10.3200/JRLP.138.5.421-432

Windschitl, P. D., Smith, A. R., Rose, J. P., & Krizan, Z. (2010). The desirability bias in predictions: Going optimistic without leaving realism. *Organizational Behavior and Human Decision Processes, 111,* 33–47. doi:10.1016/j.obhdp.2009.08.003

Zeelenberg, M., van Dijk, W. W., Manstead, A. S. R., & van der Pligt, J. (2000). On bad decisions and disconfirmed expectancies: The psychology of regret and disappointment. *Cognition & Emotion, 14,* 521–541.

Received November 11, 2011
Revision received June 28, 2012
Accepted July 6, 2012 ■