

How Chronic Self-Views Influence (and Mislead) Self-Assessments of Task Performance: Self-Views Shape Bottom-Up Experiences With the Task

Clayton R. Critcher and David Dunning
Cornell University

Self-assessments of task performance can draw on both top-down sources of information (preconceived notions about one's ability at the task) and bottom-up cues (one's concrete experience with the task itself). Past research has suggested that top-down self-views can mislead performance evaluations but has yet to specify the exact psychological mechanisms that produce this influence. Across 4 experiments, the authors tested the hypothesis that self-views influence performance evaluations by first shaping perceptions of bottom-up experiences with the task, which in turn inform performance evaluations. Consistent with this hypothesis, a relevant top-down belief influenced performance estimates only when learned of before, but not after, completing a task (Study 1), and measures of bottom-up experience were found to mediate the link between top-down beliefs about one's abilities and performance evaluations (Studies 2–4). Furthermore, perception of an objectively definable bottom-up cue (i.e., time it takes to solve a problem) was better predicted by a relevant self-view than the actual passage of time.

Keywords: self-views, expectations, bottom-up cues, self-assessment, performance evaluation

Despite spending more time with themselves than with any other person, people often have surprisingly poor insight into their skills and abilities (Dunning, 2005; Dunning, Heath, & Suls, 2004; Harris & Schaubroeck, 1988; Mabe & West, 1982). Nurses' estimates of their basic life-support skills do not correlate with their performance on objective skill tests (Marteau, Johnston, Wynne, & Evans, 1989). Adolescent boys' sense of their knowledge of condom use correlates very modestly with their actual knowledge (Crosby & Yarber, 2001). Doctors show similar poor self-insight into their knowledge of various disorders (Tracey, Arroll, Richmond, & Barham, 1997). Summing up the literature on the relationship between self-views and actual knowledge, Dunning (2005) stated that there tends to be "a real relationship between perception and reality, just not a very strong one" (p. 5; see also Dunning et al., 2004).

Clayton R. Critcher and David Dunning, Department of Psychology, Cornell University.

This research was supported in part by a National Science Foundation Graduate Research Fellowship awarded to Clayton R. Critcher, National Institute of Mental Health Grant RO1 56072, and National Science Foundation Grant 0745806, both awarded to David Dunning. We thank Jane Risen for her comments. We thank Sally Apuzzo, Claire Chung, Latonia Coryatt, Adi Kochavi, Sarah Mayefsky, Ryan Middleton, Nurul Mohamed, Bernadette Park, and Jeanette Zambito for assistance with data collection.

Correspondence concerning this article should be addressed to Clayton R. Critcher, Department of Psychology, Cornell University, 211 Uris Hall, Ithaca, NY 14853. E-mail: crc32@cornell.edu

Recently, Ehrlinger and Dunning (2003) offered one reason for why self-assessments tend to be so modestly correlated with the reality of performance. Estimates of performance are driven, at least in part, by something else—by chronic self-views people have about their abilities, preconceived notions about whether they are skilled or unskilled at a task. If John, for example, thinks that he has a good deal of logical reasoning ability, but Marvin does not, John will assume he did better on the logical reasoning quiz they both just completed than will Marvin, even if both performed equally. Ehrlinger and Dunning discovered that these preconceived notions of self often correlate equally strongly, and sometimes more strongly, with people's performance estimates than do their actual performances.

To be sure, one might think this would be a useful way to estimate one's performance. What better way to estimate one's performance in today's task than to refer to one's beliefs about one's skill, based on a lifetime of experience, in the relevant domain? The fly in the ointment of this analysis, however, is that these chronic self-views tend to be only modestly correlated with objective performance (for reviews, see Dunning, 2005; Dunning et al., 2004; Mabe & West, 1982). Thus, although these views provide a modicum of validity to performance estimates, relying on them too much can lead to bias and error (Ehrlinger & Dunning, 2003).

Ehrlinger and Dunning (2003) provided several demonstrations that people's chronic self-views influence performance assessments, irrespective of actual performance. Performance estimates

on a logic quiz were equally or more correlated, depending on the measure, with self-views than with actual performance. Changing which self-view was putatively relevant to a task caused people to provide different performance estimates, although actual performance levels were unchanged. Inducing people to have more positive or negative self-views about their knowledge of North American geography prompted them to change their performance estimates on a quiz that asked them to place famous cities (e.g., St. Louis, Missouri) on a North American map, although actual performance was unaffected. Men and women, who walked into an experiment with divergent ideas about their scientific talent, differed in how well they thought they did on a pop quiz on science, although they did not differ in objective performance. Taken together, these findings are consistent with prior research showing that more global self-views (i.e., self-esteem) predict how people evaluate their performance on specific tasks (Jussim, Coleman, & Nassau, 1987; Lindeman, Sundvik, & Rouhiainen, 1995; Shrauger & Terbovic, 1976).

But in all these demonstrations, Ehrlinger and Dunning (2003) left unanswered two important questions. The first is how do these preconceived self-notions influence performance estimates? What are the psychological mechanisms that link chronic self-views to estimates of today's specific performance? What psychological processes or mechanisms lie in the path between self-views and performance estimates?

The second mystery is why the impact of top-down self-views is so strong, given other strong influences in the environment that should dampen or eliminate that impact. In particular, it is well known in the psychological literature that people's estimates of performance are importantly driven by the bottom-up experiences they have as they complete a task. People are more confident in their performance, for example, if they come to answers quickly, rather than slowly (Benjamin, & Bjork, 1996; Costermans, Lories, & Ansay, 1992; Kelley & Lindsay, 1993; Reber & Schwarz, 1999; Schwarz, 2004; Schwarz, Sanna, Skurnik, & Yoon, 2007). They are also more confident if the terms appearing in questions and answers feel familiar to them rather than novel (Arkes, Boehm, & Xu, 1991; Schwartz & Metcalfe, 1992). Why do these bottom-up experiences fail to reduce the impact of top-down self-beliefs?

In this article, we address these two mysteries by examining the relationship between top-down self-beliefs, bottom-up experiences, and performance estimates. We suggest that bottom-up experiences may actually provide the link between self-views and performance estimates. Ehrlinger and Dunning (2003) stated, "Performance evaluations on specific tasks may not be so much bottom-up as they are top-down, formed by referring to a person's chronic view about his or her abilities in the specific domain in question" (p. 6). In essence, Ehrlinger and Dunning speculated that bottom-up cues and top-down self-views independently contribute to performance evaluation and that bottom-up evidence is set aside, in part, in favor of top-down notions. We, instead, assert that the influence of top-down and bottom-up cues may not be independent. Rather, top-down self-views influence performance estimates because they first influence people's experience and interpretation of the bottom-up cues they encounter as they complete a task, which then, in turn, informs their estimates about how well they are doing. Bottom-up cues do not dilute the impact of top-down self-views. Instead, they are the mediator responsible for the influence of those top-down views. For example, if people think

they are skilled at logic, they will perceive the amount of time they took to solve a logical brainteaser to be shorter than those who think they are unskilled. We assert that these bottom-up experiences are not different in reality, but merely in perception and interpretation. People who think they are skilled perceive themselves as having an easier bottom-up experience with the task, regardless of what objective measures may indicate.

There are many empirical demonstrations that top-down beliefs influence the phenomenology of bottom-up experience. Yogurt labeled as "full-fat" is rated as tastier than identical yogurt labeled "low-fat" (Wardle & Solomons, 1994; see also Sanford, Fay, Stewart, & Moxey, 2002). Meat labeled 75% lean tastes better than meat with the semantically identical label of 25% fat (Levin & Gaeth, 1988). A bottle of wine seems more pleasant, and activates the medial orbitofrontal cortex more, when it is priced at \$90 than at \$10 (Plassman, O'Doherty, Shiv, & Rangel, 2008). Attitudes toward samples of cola (McClure et al., 2004), turkey (Makens, 1965), seltzer water (Nevid, 1981), and beer (Allison & Uhl, 1964) were all assimilated toward the attitudes held toward these products' brands, but only when the label was showing. Moving beyond demonstrations of mere liking, Wansink, Park, Sonka, and Morganosky (2000) found that a protein bar labeled as "soy" begins to taste more grainy and less flavorful. Furthermore, bitter coffee tastes milder if tasters are first misinformed that the coffee is not actually bitter (Olson & Dover, 1978). Such effects arise at the social level, too. People who dispositionally expect to see interpersonal hierarchies were more likely to see a hierarchy difference between two photographed people (Mast, 2005). Labeling a face as African American causes people to see its skin as darker than when the face is labeled as European American (Levin & Banaji, 2006). A push looks more aggressive if it comes from an African American rather than a European American protagonist (Sagar & Schofield, 1980).

Here, we assert that people's top-down self-beliefs similarly color the phenomenology and interpretation of bottom-up cues while completing a task. To our knowledge, there have been no empirical demonstrations of whether top-down self-beliefs influence people's *in vivo* bottom-up experience and whether that influence, in turn, shapes estimates of performance. The closest study to have looked at such phenomena is that of Bunz, Curry, and Voon (2007), who did not examine experience with an actual task but did find that "perceived fluency" (i.e., perceived skill) at computer tasks correlated with a report that one had "computer-related anxiety." Furthermore, this anxiety was not related to actual ability. Bunz et al.'s purpose was not to measure bottom-up experience, and in fact, the items on their anxiety scale seem to measure a sense of ability (e.g., "I feel insecure about my ability to interpret a computer printout"), rather than a bottom-up experience. Furthermore, this computer-related anxiety reflected a general, decontextualized belief, not an assessment of actual experience with a computer task. Instead, we propose (and test) that top-down beliefs color *in vivo* bottom-up experience. Also, Bunz et al.'s study did not examine the second part of our hypothesis, that these contaminated bottom-up experiences mediate the link between self-views and performance estimates on a specific task. Herein, we expressly focused on whether chronic self-views influenced one's subjective experience as one sits down to actually perform a task—and whether those altered experiences ultimately led to different notions about how well one had done on a task.

Overview of the Present Studies

In four studies, we tested whether chronic self-views influence performance evaluation because they influence the perceptions and interpretations of the bottom-up experience people have with the task. Study 1 explored this analysis by varying the timing of when participants learned that a certain skill (e.g., computer programming) was supposedly relevant to the task at hand. We hypothesized that if participants learned about the skill before beginning the task, then their self-beliefs of skill would color bottom-up experiences and, thus, influence performance estimates. However, if they learned about the skill after completing the task but before providing performance evaluations, there would be no impact of top-down self-views on performance estimates, given that their bottom-up experiences had already been set. Study 2 provided a more direct test by actually testing statistically whether bottom-up experience mediates the link between self-views and performance estimation. Studies 3 and 4 did the same but added more direct tests of mediation by manipulating which self-view was supposedly relevant to the task at hand.

Study 1

The goal of Study 1 was to provide an initial test of the idea that top-down self-views influence performance evaluations through their impact on bottom-up experience. The study extended and replicated Ehrlinger and Dunning (2003, Study 2), in which participants were told that a Graduate Record Examination (GRE)-type test measured either abstract reasoning (a skill participants tended to believe that they had in abundance) or computer-programming skill (a skill participants did not think they possessed to any positive degree). Ehrlinger and Dunning found that participants provided more favorable evaluations of their own performance (e.g., thought they got more questions right) when they thought it measured abstract reasoning, rather than computer skill, although there was no difference in actual performance.

Our extension involved varying exactly when participants were told that the test focused on abstract reasoning versus computer-programming skills. Some participants were informed before they took the test, with the remainder being told after they had completed the test but before they evaluated their performance. If top-down views influence performance estimates directly, then the timing should not matter. However, because we suggest that top-down views influence performance estimates indirectly through participants' bottom-up experiences, we predicted that the timing would matter. Participants told before the test had a chance to have their self-views influence their bottom-up experiences with the test and, thus, their performance estimates. Those informed afterward would already have concluded their bottom-up experiences with the test, and thus, their performance evaluations should remain unaffected.

There is empirical precedent for varying the timing of providing top-down information to test whether top-down beliefs exert their influence by acting through bottom-up perceptions. Learning that balsamic vinegar had been added to a sample of beer (which apparently sounds much worse than it tastes) contaminates one's enjoyment only when learning of this mystery ingredient before sampling it, not afterward (Lee, Frederick, & Ariely, 2006). More generally, when top-down views influence perceptual (bottom-up)

processing, the biasing view must be in place at the time of exposure to the bottom-up, perceptual information (see von Hippel, Sekaquaptewa, & Vargas, 1995).

Method

Participants and Design

Forty-five undergraduates at Cornell University (Ithaca, New York) participated in exchange for course extra credit. Participants were randomly assigned to one of the cells in a 2 (Timing: pretest or posttest) \times 2 (Skill Related: abstract reasoning vs. computer programming) full-factorial design.

Procedure

As participants arrived at the laboratory, we first assessed their self-views in a number of domains, including abstract reasoning and computer programming but also filler domains (e.g., emotional intelligence, geography). Participants indicated their skill level on percentile scales, ranging from 0 to 100, that asked them to compare their skill with those of other students at Cornell University by noting the percentage of peers they thought they outperformed for each skill. Then, participants indicated for the same domains "the extent to which you believe you possess the following abilities or aptitudes" on 11-point Likert scales, anchored at 1 (*not at all*) and 11 (*completely*).

At this point, participants in the pretest timing condition learned that they would be taking a test of abstract reasoning or computer-programming ability. Those in the posttest timing condition were simply told they would be taking an "ability test." At this point, participants took the 10-item logic games test used in Ehrlinger and Dunning (2003, Study 2). At the end of the test, participants in the posttest condition learned that they had just taken a test of abstract reasoning or computer-programming ability. Finally, participants estimated how many of the 10 items they believed they answered correctly and at what percentile they believed their specific performance fell relative to other students taking part in the experiment.

The zero-order correlations between performance and the measured variables (for all four studies) are listed in Table 1.

Results and Discussion

As a check that self-views were more positive for abstract reasoning than for computer-programming ability, we conducted two 2 (Timing: pretest or posttest) \times 2 (Relevant skill: abstract reasoning vs. computer programming) \times 2 (Skill Rated: abstract reasoning vs. computer programming) mixed-model analyses of variance, with the last factor measured within-subjects. For the percentile scores, participants saw their own abstract reasoning ability skills ($M = 64.7$, $SE = 2.4$) as superior to their computer-programming skills ($M = 41.2$, $SE = 4.9$), $F(1, 41) = 26.62$, $p < .001$. On Likert-scale responses, abstract reasoning skills ($M = 7.2$, $SE = 0.29$) were also seen to be superior to computer programming skills ($M = 4.7$, $SE = 0.46$), $F(1, 41) = 32.01$, $p < .001$. None of the higher order effects involving condition approached significance (all F s < 1), speaking to the success of random assignment.

Table 1
Correlations Between Actual Performance, Relevant Self-View, Bottom-Up Experience, and Estimated Performance, by Study

Study	Self-view	Bottom-up experience	Estimated performance
Study 1			
Actual performance	.43**		.66***
Self-view			.37*
Study 2			
Actual performance	.03	.13	.15*
Self-view		.16*	.15*
Bottom-up experience			.39***
Study 3			
Actual performance	.09	.53***	.57***
Self-view		.26**	.37**
Bottom-up experience			.65***
Study 4: GRE test			
Actual performance	.27***	.33***	.24**
Self-view		.32***	.37***
Bottom-up experience			.67***
Study 4: High school test			
Actual performance	.30***	.35***	.31***
Self-view		.39***	.43***
Bottom-up experience			.68***

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

For the performance evaluations taken after completion of the test, both raw score and percentile estimates were highly correlated, $r(42) = .71$, $p < .001$. One participant's performance estimates fell more than seven standard deviations from the condition mean; this outlier was excluded from all analyses reported below. At this point, we standardized the two performance-estimate items and summed them to create a performance-estimate composite and submitted this composite to a two-way Timing \times Relevant Skill analysis of covariance, with test label and timing as between-subjects variables and actual test performance as a covariate.¹ The Timing \times Skill interaction was significant, $F(1, 38) = 4.41$, $p = .04$. As seen in Figure 1, the significant interaction reflected the tendency for the relevant skill to affect performance estimates only when participants learned of it before taking the test, $t(38) = 3.08$, $p = .004$, but for it to have no effect when learned after the test ($t < 1$).

We then tested whether individual differences in participants' self-views could predict performance estimates when participants learned of the test label before the test but not after the test. Thus, we regressed the performance composite estimate on the timing condition, the relevant self-view (as measured by self-rating), the irrelevant self-view (again, the self-rating), the Relevant Self-View \times Timing interaction, the Irrelevant Self-View \times Timing interaction, and the actual score. As predicted, the Relevant Self-view \times Timing interaction reached significance, $\beta = -.31$, $t(36) = 2.67$, $p = .01$, whereas the Irrelevant Self-View \times Timing interaction did not ($\beta = .10$, $t < 1$). We conducted simple slopes analyses to examine this significant interaction (Aiken & West, 1991). Consistent with hypotheses, performance estimates were tightly correlated with self-views in the pretest condition, $\beta = .73$, $t(36) = 3.39$, $p = .002$, and showed no relationship in the posttest condition ($\beta = .08$, $t < 1$).² This shows that the key interaction emerges not merely using a dichotomous coding for the relevant self-view but using an idiographically assessed measure as well.

Finally, to confirm that the effect of test label on performance estimates (in the pretest condition) could be entirely explained by the idiosyncratic self-views, we regressed performance estimates on the test-label condition and the relevant self-view, while controlling for actual performance. As expected, the relevant self-view predicted performance estimates, $\beta = .61$, $t(16) = 3.46$, $p = .003$, whereas the effect of test label became nonsignificant, $\beta = .19$, $t(16) = 1.21$, $p > .24$. In other words, differences in perceived skill at abstract reasoning and computer programming explained the gap between the two pretest conditions.

In sum, self-views influenced performance estimates, but only when participants learned what self-view was relevant before taking the test. These results are consistent with our hypothesis that top-down views only contaminated performance estimates when they had the potential to distort one's bottom-up experience. When the relevant self-view was learned after the test, it no longer could contaminate one's bottom-up experience with the test. As expected, making different self-views relevant to the test left performance estimates unaffected.

These results may seem somewhat surprising in light of research on the *perseverance effect* (Ross, Lepper, & Hubbard, 1975), the tendency for task feedback to continue to influence performance estimates for the future, even after the feedback has been discredited. Feedback in perseverance studies is given after, not before, the person completes the task (e.g., Guenther & Alicke, 2008; McFarland, Cheam, & Buehler, 2007), leading to an apparent discrepancy with the present study. That is, perseverance researchers see an effect for post-task information, whereas we do not.

We would point to a key difference between our work and the research done on perseverance that may explain the different findings for post-task feedback. In their feedback, perseverance researchers give their participants a specific and pointed event to explain. Participants are told that they either succeeded brilliantly at a task or that they failed. Perseverance theories like that of Ross et al. (1975) have argued that this false performance feedback prompts people to construct causal stories to explain their performance—and that these performance-explaining explanations re-

¹ By covarying out actual performance, we could examine the influence of top-down or bottom-up cues that were not simply sensitive to true performance. In other words, we were not interested in the pathway by which accurate self-views influenced performance estimates through bottom-up experience because one's bottom-up cues detected one's actual performance.

² The astute reader may have noticed that those who learned of the relevant self-view after taking the test gave performance estimates similar to those who learned that the test was one of abstract reasoning before taking the test. Given that the test actually was one of abstract reasoning abilities (logical reasoning), this pattern is perhaps unsurprising. But to further confirm that those in the after-test condition were spontaneously drawing upon their abstract reasoning self-views in completing the test, we regressed these after-test participants' performance estimates on their abstract reasoning and computer-programming self-views, while controlling for actual performance. Consistent with our assumption that participants in the posttest condition spontaneously labeled the test as one of abstract reasoning, there was an almost-significant influence of abstract-reasoning self-view on performance estimate, $\beta = .27$, $t(19) = 2.03$, $p = .06$, whereas there was no relationship between computer-programming self-views and estimated performance ($\beta = -.10$, $t < 1$).

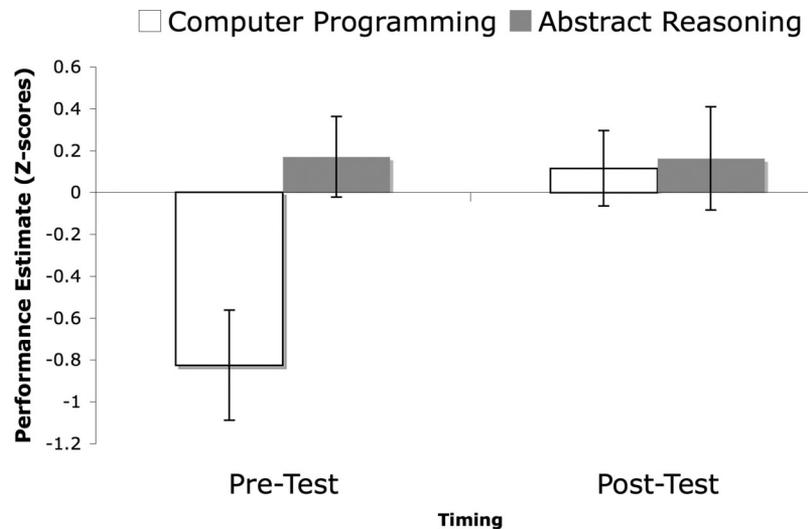


Figure 1. Performance estimates by test label and when the test label was revealed (Study 1). Error bars indicate ± 1 standard error.

main even after the feedback is discredited. In the language of top-down and bottom-up cues, performance feedback leads people to call upon other top-down theories that would explain their success or failure. That is, it does not appear that false feedback leads people to reconsider what their experience with the task was like—that is, to review their bottom-up experience—but, rather, to bring to mind top-down theories that would explain that event. Note that, for us, the post-task information given in Study 1 was different. Participants were not given an event to explain but, rather, just one top-down cue that they could use to guess what that event (i.e., their performance) was. As an example, participants in the posttest computer-programming condition may have been surprised that they did not have more difficulty with the test, learning of the tested domain did not demand that they revise their understanding of the test experience. As such, their performance estimates were not influenced.

There does remain one alternate model with which Study 1 is also compatible. It is possible that knowledge of the relevant self-view before the test did not distort one's bottom-up experience but, instead, discouraged one from carefully attending to bottom-up information about a task. This perspective is reflected in the *cognitive miser* approach (Fiske & Taylor, 1991), which holds that people rely on prior top-down beliefs so they do not have to expend the cognitive resources to attend to bottom-up cues (Belmore, 1987; Fiske, Neuberg, Beattie, & Milberg, 1987). This alternative predicts that only those in the posttest condition attended to performance-relevant bottom-up cues and that these undistorted cues eliminated the influence of the self-view manipulation. To rule out this alternative, in the remaining studies, we actually measured participants' bottom-up experience to test whether self-views did in fact distort it.

Study 2

Study 1 provided indirect evidence that top-down views have an impact on performance evaluation because they alter the ways in

which bottom-up cues are perceived. However, in Study 1, we did not directly measure bottom-up cues, and thus, it was possible that top-down views contaminated performance estimates, not because they changed bottom-up experience, but perhaps because the information the self-view putatively conveyed discouraged people from attending to bottom-up cues. Thus, in Study 2, we decided to examine directly whether the link between chronic self-views and performance assessments flowed through perceptions of bottom-up experiences.

Participants completed a 15-item DVD-based test of interpersonal perception abilities. Before the test, we asked participants to report on their interpersonal perception ability. Then, after answering each question on the test, participants answered five further queries tapping into their bottom-up experience. We examined whether self-views would be correlated with these bottom-up experiences, controlling for actual performance, and whether these bottom-up perceptions would explain any relationship between self-views and summary performance evaluations that participants provided after taking the test. In other words, we tested whether the influence of self-views on performance estimates could be statistically mediated by bottom-up experience.

Method

Participant

Two hundred forty-two members of the Cornell University community participated in exchange for course extra credit. Participants completed the study as part of an hour-long session in which they also completed an unrelated experiment.

Procedure

First, all participants stated in what percentile among students at their university they believed they fell (from 0th to 100th) on several abilities. It was crucial that participants indicated their

percentile placement for *social perception ability*, which was defined for participants as “the ability to accurately observe and interpret the expressive behavior of others, to decode behavioral and verbal cues from others in order to reach accurate assessments about them, and to interpret subtle expressive behaviors.” Participants wrote their percentile estimate in a blank provided.

Participants then learned that they would be taking the Interpersonal Perception Task (IPT), “a measure that indicates accuracy in social perception.” The IPT-15 (Costanzo & Archer, 1989) presents 15 audiovisual scenes of interpersonal interaction. The test taker must use verbal and nonverbal cues to discern the relationships between the characters. For example, in one scene, a man and a woman have a short discussion about teaching at a university. At the end of the scene, the test-taker must indicate, on the basis of what he or she observed, which person is the higher status person.

After participants indicated the answer to each question, we had them answer five questions designed to tap into their bottom-up experiences. On 5-point scales, participants indicated how long it took them to figure out the answer, whether they immediately knew the answer, to what extent they had to guess, the perceived difficulty of the item, and the extent to which they went back and forth between the answer choices. Finally, at the end of the test, participants estimated how many of the 15 items they believed that they answered correctly and in what percentile they believed that their performance placed them.

Results and Discussion

As expected, participants’ estimated test score and estimated percentile score on the test were correlated, $r(240) = .45$, $p < .001$. Accordingly, we standardized each and summed these values to create a performance-estimate composite. We then conducted four analyses.

First, to test whether we replicated Ehrlinger and Dunning (2003), we regressed the performance-estimate composite on ratings of social perception ability that participants provided before taking the test, controlling for participants’ actual test score. Although actual performance predicted performance estimates, $\beta = .14$, $t(239) = 2.24$, $p = .03$, participants’ self-views were especially predictive, $\beta = .27$, $t(239) = 4.38$, $p < .001$.

Second, to assess whether participants’ self-views influenced bottom-up experiences with the test, we first conducted a principal-components analysis on the five bottom-up queries for each of the 15 test items separately. In each case, a single-factor solution emerged. Thus, we took each participant’s factor score for each of the 15 items and averaged them together to create a bottom-up composite of bottom-up experience for each participant.³ Higher numbers indicated that the item was experienced as easier. We regressed this bottom-up composite on self-views while controlling for actual performance. Participants’ actual performance was only marginally predictive of their bottom-up experience, $\beta = .12$, $t(237) = 1.92$, $p = .06$, but their self-views significantly predicted their bottom-up experience with the items, $\beta = .15$, $t(237) = 2.43$, $p = .02$.

Third, we tested whether bottom-up experiences were predictive of participants’ performance estimates. We regressed performance estimates on bottom-up experiences while controlling for actual performance. Again, actual performance was only marginally predictive of performance estimates, $\beta = .11$, $t(237) = 1.92$, $p = .06$,

but bottom-up experiences closely corresponded with participants’ performance estimates, $\beta = .39$, $t(237) = 6.51$, $p < .001$.

Finally, to examine whether bottom-up experiences mediated the link between chronic self-views and performance estimates, we tested whether bottom-up experiences would continue to predict performance estimates after we controlled for self-views (Baron and Kenny, 1986). Bottom-up experiences continued to predict performance estimates, $\beta = .35$, $t(237) = 6.02$, $p < .001$. In addition, the direct effect of self-views on performance estimates was significantly reduced, as indicated by a Sobel test ($z = 2.25$, $p = .02$), although the effect of self-views on performance estimates remained significant, $\beta = .22$, $t(237) = 3.73$, $p < .001$. Taken together, these tests suggested that bottom-up experience partially mediated the link between self-views and performance estimates (see Figure 2).

We should note that participants’ self-views did not correlate with their actual performance, $r(240) = .03$, *ns*. Thus, to the extent that self-views contaminated bottom-up cues, the self-views led performance estimates astray. However, bottom-up experience did correlate, albeit weakly, with participants’ actual performance, $r(240) = .13$, $p = .05$.

In sum, even when we controlled for their actual performance, participants’ performance estimates on a test of social perception ability was predicted by their self-views of their social perception ability, with this relationship being partially mediated by their ratings of bottom-up cues. Although self-views exerted a top-down influence on self-evaluation, they appear to do so partially by their impact on bottom-up experience. Of course, this study was correlational, so it would be a stronger test of our hypothesis to show that manipulating self-views could influence participants’ performance estimates by altering the bottom-up experiences they believe they had while completing a task.

Study 3

Study 3 was designed with two goals in mind. First, we returned to the methods of Ehrlinger and Dunning (2003), using a test that we could reasonably claim measured either abstract reasoning or computer-programming ability, and this time, we asked participants to report their bottom-up experiences as they completed the test. We expected that participants’ bottom-up experience would be driven by the self-view we said was relevant to the test, but not the self-view that was irrelevant, and that it would be the relevant self-view that would again predict performance evaluation. This would demonstrate that the manipulated self-view changed participants’ bottom-up experience.

Second, we wanted to test whether the top-down beliefs changed only the subjective bottom-up experience with the task, rather than objective bottom-up experience. Although participants were not aware of this, the computer timed how long it took them to

³ For all bottom-up items except one (*knew immediately*) higher numbers tended to indicate more trouble with the question. Many participants seemed to miss this difference, giving uniformly high or low numbers to all five bottom-up items for each test question. By using the factor scores instead of simply averaging the bottom-up responses together, this “noisy” item was deweighted, given that it, of course, loaded less highly on the single factor. Factor scores are used for the same reason in the analyses for Study 3 as well.

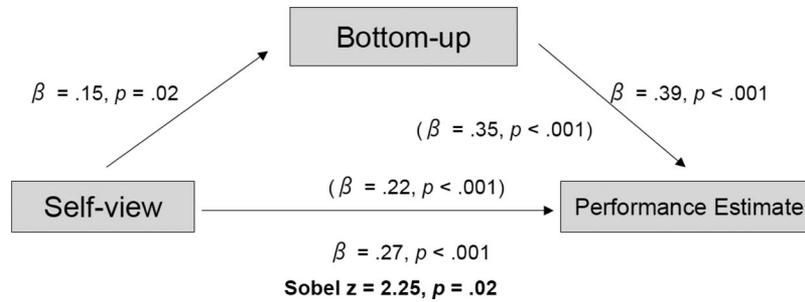


Figure 2. The mediation model demonstrating that bottom-up experience partially mediates the relationship between chronic self-views and performance estimates (Study 2). Note that the standardized betas in parentheses are those from the regression model in which performance estimate was regressed on self-view and bottom-up experience simultaneously.

complete each question. In this way, we could see if the expectation that a task would be easier actually led participants to speed through it more quickly, or whether self-views only produced the *perception* that they were flying through the task, even if that perception had no relationship to actual experience.

Method

Participants

One hundred thirty-three undergraduates at Cornell University participated in the study in exchange for extra course credit.

Procedure

Participants first rated four of their abilities on 11-point Likert scales, which included ratings of their abstract reasoning ability and computer-programming ability on scales anchored at 1 (*not at all*) and 11 (*completely*). Then, depending on condition, participants were told they would be taking a test that tapped into abstract reasoning ability or computer-programming aptitude. Participants completed the same 10-item test used in Study 1, although items were presented one at a time on a computer, rather than on paper.

After completing each item, participants answered five questions designed to assess their bottom-up experience. Because the format of the test was different from Study 2, some of the bottom-up cues seemed less relevant (e.g., going back and forth between answer choices). Instead, participants indicated the extent to which they questioned the logic they used; how long it took them to solve a problem; the extent to which it felt like they were guessing; the extent to which the way to solve the problem came to them immediately, leading them to answer the question quickly; and the perceived difficulty of the question. After completing the test, participants estimated how many of the 10 items they answered correctly and in what percentile they believed their performance fell.

Results and Discussion

As expected, participants rated their abstract reasoning ability more favorably ($M = 6.9$) than their computer-programming ability ($M = 3.3$), paired $t(132) = 16.93, p < .001$. As in Study 1, participants' percentile estimates and score estimates were posi-

tively correlated, $r(131) = .45, p < .001$. We once again standardized and summed these values to create a performance-estimate composite.

To assess whether the self-view manipulation led to between-condition differences in performance assessments, we submitted participants' performance estimates to a one-way analysis of covariance with actual performance as a covariate. Participants who believed they were taking the abstract reasoning test believed they had performed better ($M = 0.32, SE = 0.18$) than those who completed the computer programming test ($M = -.24, SE = 0.16$), $F(1, 130) = 5.38, p = .02$, even though participants' actual performance did not differ between the abstract ($M = 7.6, SD = 1.84$) and computer programming ($M = 7.7, SD = 1.97$) conditions ($F < 1$). Contrary to our expectations, bottom-up experience did not differ by condition ($M_s = -.01$ and $.01$), $F < 1$. Nonetheless, our remaining analyses suggest our manipulation did indeed change bottom-up experience, and we return to this oddity in Study 4 with a methodological modification.

Mediation Analyses

To examine whether the relevant, but not the irrelevant, self-view predicted performance estimates through an effect on bottom-up experience (despite the lack of a main effect of condition on bottom-up experience), we conducted mediational analyses for each condition. All regression analyses controlled for actual test performance, so we concentrated on influences on bottom-up experience and performance estimates that did not stem from actual performance. As in Study 2, we combined the responses to the bottom-up questions for each pair of items using principal-components analysis. For each participant, we averaged across these five factor scores to create a bottom-up composite for each participant. Higher scores indicated a more positive bottom-up experience with the test. A summary of analyses is presented in Figure 3.

Abstract reasoning test. We regressed participants' estimated performance on the abstract reasoning and computer-programming self-views. As expected, participants' abstract reasoning self-views predicted performance estimates, $\beta = .25, t(54) = 2.15, p = .04$, whereas their computer-programming self-views did not relate to performance estimates ($t < 1$). We then regressed bottom-up experience on the two self-views. Again, abstract reasoning self-

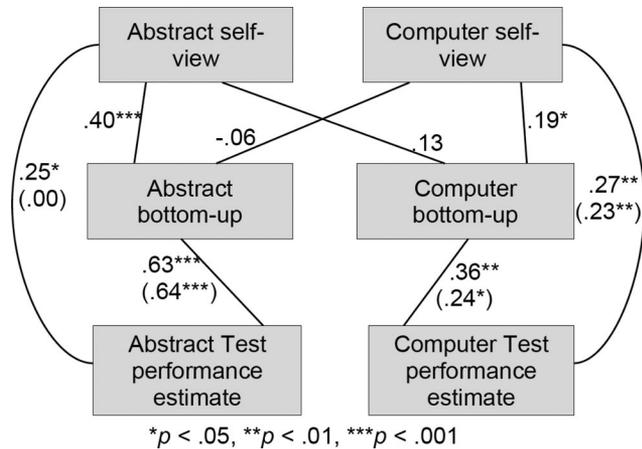


Figure 3. The mediation models demonstrating that bottom-up experience mediated the relationship between participants' abstract reasoning and computer-programming self-views on their performance estimates on what was labeled an abstract reasoning and computer-programming test, respectively (Study 3). Note that all numbers are standardized betas. The betas in parentheses are from the regression model in which performance estimate was regressed on the test-appropriate self-view and bottom-up experience simultaneously.

views predicted bottom-up experience with the test, $\beta = .40$, $t(54) = 3.48$, $p < .001$, whereas computer-programming self-views' influence was nonsignificant ($t < 1$). Bottom-up experience also drove participants' performance estimates, $\beta = .63$, $t(55) = 6.50$, $p < .001$. When regressing performance estimates on the bottom-up experience and both self-views, the bottom-up experience remained a strong predictor of performance estimates, $\beta = .64$, $t(53) = 5.86$, $p < .001$, whereas both self-views no longer offered any explanatory power in predicting performance estimates ($t_s < 1$). A significant Sobel test ($z = 2.02$, $p = .04$), suggested that the influence of abstract self-views on test performance was fully mediated by the self-views' influence on bottom-up experience.

Computer-programming test. For participants told the test tapped into computer-programming ability, computer-programming self-views predicted performance estimates, $\beta = .27$, $t(71) = 3.33$, $p = .001$, although abstract reasoning self-views did as well, $\beta = .19$, $t(71) = 2.31$, $p = .02$. However, in looking at the influence of self-views on bottom-up experience, only computer-programming self-views affected bottom-up experience, $\beta = .19$, $t(71) = 2.02$, $p = .05$, whereas abstract reasoning self-views did not, $\beta = .13$, $t(71) = 1.46$, $p > .14$. Bottom-up experience, in turn, was a predictor of performance estimates, $\beta = .36$, $t(71) = 3.42$, $p = .001$. When we regressed performance estimates on bottom-up experience and both self-views, bottom-up experience remained a significant predictor of performance estimate, $\beta = .24$, $t(70) = 2.37$, $p = .02$, as did the computer programming self-view, $\beta = .23$, $t(70) = 2.79$, $p = .01$. The abstract reasoning self-view dropped to marginal significance, $\beta = .16$, $t(70) = 1.94$, $p = .06$. Because the abstract reasoning self-view did not influence bottom-up experience, bottom-up experience was only a candidate mediator for the relationship between computer-programming self-view and performance estimate. A Sobel test provided evidence for partial mediation ($z = 1.93$, $p = .05$).

Subjective Versus Objective Bottom-Up Experience

We next examined whether self-views influenced only the participants' subjective experiences with the task or whether they also influenced objective experiences. We did so by examining the experience of time. To create an index of perceived time, we averaged the five ratings (one for each pair of items) of the extent to which participants solved the problem quickly (reverse scored) with the five items asking how long it took participants to solve the problem.⁴ This composite had good internal reliability ($\alpha = .75$), especially given the fact that we aggregated across bottom-up experience from different items. As a measure of actual time, we examined the average number of milliseconds that participants took from the time the multiple-choice question appeared on the screen to the point at which they indicated a response. Because, across participants, there was evidence of positive skew, we log transformed these times. We then regressed the perceived time index on both self-views and the log-transformed actual times.

For those in the abstract reasoning condition, the abstract reasoning self-view predicted a sense that participants were taking less time to complete the items, $\beta = -.46$, $t(54) = 3.87$, $p < .001$, whereas the computer-programming self-view had no effect ($t < 1$). The actual time participants spent on the question did not relate to how long they felt they were taking, $\beta = .12$, $t(54) = 1.01$, $p > .31$. For those in the computer-programming condition, more positive computer-programming self-views led participants to believe they were completing the items more quickly, $\beta = -.22$, $t(71) = 1.97$, $p = .05$, whereas the abstract reasoning self-view had a marginal influence, $\beta = -.19$, $t(71) = 1.70$, $p = .09$. Again, the actual time participants spent on the items did not predict their sense of how long it was taking them, $\beta = .18$, $t(71) = 1.61$, $p > .11$. Even collapsing across conditions, there was no significant correlation between the log-transformed actual times and the perceived time, $r(131) = .13$, $p > .12$. Furthermore, there was no tendency for those with more positive self-views to actually complete the items any more quickly; if anything, they completed the items more slowly $r(131) = .10$, $p > .23$.

In sum, despite taking the same test, participants' performance evaluations corresponded to their self-rated views of skill regarding abstract reasoning ability or computer-programming ability, depending on how we labeled the test. The relationships we observed between self-ratings of skill and performance estimates were mediated by the influence of the appropriate self-view on bottom-up experience, which then connected to performance estimates. Also, participants' perceptions of the time did not correspond to the actual passage of time. Instead, subjective time perception was predicted by the self-view purportedly relevant to the test.

⁴ To determine the reasonableness of this composite, we presented 60 participants with the five original items and asked them to select the item or items that essentially asked about "the perceived length of time it took to solve the problem" and to rate each item on a 1–4 scale for whether it should be included in a perceived time composite. A majority selected these two items. A majority rejected the other three items. On the continuous measure, only these two items had means greater than 3. Regardless of whether one looks to the dichotomous or continuous measure, the two items making up our composite were judged superior to the other three items ($p_s < .001$).

Given that bottom-up experience was differentially tied to different self-views in each condition, it is clear that our manipulation of top-down views changed participants' bottom-up experience. For example, the only way that computer-programming self-views could be unrelated to self-views in the abstract reasoning condition but related to self-views in the computer programming condition is if the manipulation shifted participants' bottom-up experience into alignment with this self-view. Furthermore, our condition manipulation changed participants' performance estimates. Like the bottom-up experience, performance estimate was differentially tied to the two self-views depending on participants' condition. Finally, the two significant mediation models reflect that self-views influenced performance estimates by influencing bottom-up experience.

Study 4

Nonetheless, there remained one oddity in the data from Study 3: Making a different self-view relevant to the test changed performance estimates between conditions but did not produce a similar main effect for bottom-up experiences. That is, participants in the abstract reasoning condition did not rate their bottom-up experiences of the test as more benign relative to those in the computer-programming condition. How can there be clear evidence that self-views changed bottom-up experience at the individual level but no overall main effect between conditions?

One explanation for this oddity is that participants in Study 3 may have used the bottom-up scales differently by condition because they were using different reference points to inform their reports. People in the computer-programming condition, for example, may have actually experienced the test as more difficult, but then moderated their bottom-up reports to indicate that although the test was tough, it was "not *that* tough for a computer science test." This would explain why the mediation models within each condition supported our hypothesis and why there was a between-condition main effect on performance estimates but not bottom-up experience. Other work has shown that people shift their usage of rating scales depending on their expectation or experience in a way similar to what we suspect here. For example, people shift how they use a response scale about height or aggression depending on whether they are talking about a man or a woman. That is, a woman need not be as tall as a man to receive a rating of *tall* from perceivers (for a review, see Biernat, 2005), a response tendency known as the *shifting standards* phenomenon. Respondents have also been shown to shift how they use response scales depending on their cultural context. For Asian respondents to rate themselves as *respectful of elders*, they need to show more of that respect on an objective, behavioral level than would a respondent from North America (Peng, Nisbett, & Wong, 1997).

To minimize this possible artifact in Study 4, we manipulated self-views within participants, rather than between them, so that participants would be more likely to define the bottom-up scales in a consistent way across both tasks they completed. In other words, if we asked participants between subjects to rate on a 9-point scale the height of a 6 ft. 3 in. (1.9 m) man and a 5 ft. 9 in. (1.75 m) woman, we imagine that both targets would be rated an 8 or a 9. But if we asked participants to place both targets on the same scale, it would be more likely that the man would receive a higher numerical judgment than the woman.

In Study 4, participants took two 15-item tests of U.S. history. They were told that one of the tests included items from the state of Maine's high school exit exam. The other test supposedly included items from a graduate entrance exam for history PhD programs. We predicted that people would report a more confident performance estimate and easier bottom-up experience when the task was described as a high school history test (for which participants should have a favorable self-view) than when it was described as a graduate-level history test (for which participants would have a more negative self-view). Further, we expected that a within-subjects mediation analysis (Judd, Kenny, & McClelland, 2001) would indicate that differences in bottom-up experience explained the link between self-view and performance assessments.

Finally, we should mention that we altered the focus in our measures of bottom-up experience. Bottom-up experience questions related to whether one is confident in one's logic in working through a problem are appropriate for tests of logical reasoning but are less so for tests of history. As we have mentioned above, research shows that people are more confident in their responses to the extent that they find test material to feel familiar, rather than novel (Arkes et al., 1991; Schwartz & Metcalfe, 1992). Therefore, our bottom-up experience questions dealt with the perceived familiarity and fluency of the material.

Method

Participants

One hundred fifty-nine undergraduates at Cornell University participated in exchange for extra course credit.

Procedure

Participants were told they would be completing two tests of U.S. History that contained past questions from the GRE: History and the Maine High School Exit Exam. To lend validity to our cover story, we asked participants to indicate whether they had taken either test between 1994 and 1998, the years from which the questions were supposedly taken. In addition, we explained that during that interval, one of the tests had shifted from four to five answer choices, which explained why one of the tests had a variable number of answer choices. We hoped that this small difference would lend credence to our cover story that the tests indeed came from different sources. Each test was in a separate test booklet, and the cover pages contained the official seal of the GRE and the Educational Testing Service or of the state of Maine. In addition, the text "Must return to experimenter at end of session. Officially approved for in-lab use only!" appeared in a box at the bottom of both cover pages.

The order in which participants completed the tests was counterbalanced. We also used two tests. For roughly half of the participants, Version A was described as the high school-level test and Version B the graduate school version. For the remainder, the pairing of version with its label (high school vs. graduate school) was reversed. As such, participants overall confronted tests of equal difficulty across high school and graduate school conditions.

Before beginning the test, participants indicated on 1 (*not at all*) to 11 (*completely*) scales the extent to which they knew U.S. history as would be required on a high school exit exam and on a

graduate school entrance exam. Participants went through each 15-item test, writing their multiple-choice answers on a provided page. After every three questions, participants were to indicate their bottom-up experiences from each of the previous three questions. Those items included: knew the answer immediately upon reading the answer choices; regardless of how I know, the answer feels like the right one; and I have a specific memory of having been exposed to this information before. Participants indicated their degree of agreement with each item on 4-point scales, with higher numbers indicating greater agreement.

After completing both tests, the experimenter collected the answers and bottom-up cues and provided participants with a follow-up sheet. Participants first indicated how many items out of 15 they believed they answered correctly on both tests. Then, participants completed four questions that asked them to retrospectively evaluate their bottom-up experience on each test. Participants indicated their confidence in their answers to the questions, the extent to which it felt like the questions dealt with content they had learned before, to what extent they marked items that were backed up with things they knew or had once learned, and how often it felt like their responses were simply random guesses.

Results and Discussion

Indicating that the manipulation was successful, participants had more confidence in their high school U.S. history abilities ($M = 7.1$, $SD = 2.1$) than in their graduate entrance exam U.S. history abilities ($M = 3.6$, $SD = 1.9$), paired $t(158) = 22.50$, $p < .001$.

We begin below by presenting the results of our within-subjects analytic approach. We then also tested our hypotheses using a between-subjects approach that is more consistent with our earlier studies by looking only at the test that participants took first. Even though the between-subjects approach would not correct for a shifting-standards problem, and thus would be expected to produce weaker results, it is possible that we would see less of a shifting-standards problem than in Study 3, given that all participants knew while taking the first test that they would be taking both a high school and a GRE test and that they would be making their bottom-up ratings on the same scale.

Within-Subjects Analyses

We first tested whether self-views influenced participants' performance estimates independent of any actual performance differences. Controlling for which test version was paired with which test label, participants performed no better on the high school test ($M = 6.5$, $SE = 0.16$) than on the GRE test ($M = 6.7$, $SE = 0.17$), $F(1, 157) = 1.74$, $p > .18$, suggesting that the test labels themselves did not systematically influence actual performance. However, controlling for participants' actual performance and version of the test/label pairing, participants estimated that they had answered more items correctly on the high school test (adjusted $M = 5.8$, $SE = 0.20$) than the GRE version (adjusted $M = 4.9$, $SE = 0.20$), $F(1, 156) = 30.53$, $p < .001$.

To determine whether the self-view influenced bottom-up experience with the task, we began by averaging together the bottom-up ratings for each question.⁵ We then had a composite for each of the three cues—knew immediately; regardless of how I know, it just felt right; felt like I had learned this before—for each

of the tests. For each comparison, we included both the test-version label pairing and participants' actual performance as covariates. While taking the high school test, participants' bottom-up experience with the task suggested they experienced the task as easier, $F(1, 153) = 18.16$, $p < .001$.

We have shown that participants had more positive self-views regarding the high school test compared with the GRE, reported having a more positive bottom-up experience with the high school test, and then incorrectly perceived that they had performed better on the high school test. To demonstrate within-subject mediation—that the influence of self-view on performance estimates can be explained by different bottom-up experiences—we computed the correlation between performance estimates and bottom-up experience, while controlling for actual performance (Judd, Kenny, & McClelland, 2001). Consistent with our mediation model, online bottom-up experience strongly correlated with performance estimates for both the high school test, $pr(153) = .67$, $p < .001$, and the GRE test, $pr(153) = .64$, $p < .001$.

To further confirm the mediation model, we examined how participants responded to the high school test relative to the graduate school test along three difference scores (high school minus graduate school rating) for actual performance, perceived performance, and bottom-up cues. If the impact of the self-view manipulation is mediated by bottom up cues, then the difference score of bottom-up cue ratings should be correlated with the difference score of perceived performance (Judd et al., 2001). That was, indeed, the case (controlling for the actual score difference), $pr(153) = .67$, $p < .001$. In short, to the extent that participants rated their bottom-up experience with the high school test as easier than the graduate school test, they also perceived their performance on the high school test to be higher relative to the graduate school test.⁶

⁵ We used averaged composites, instead of factor scores (as in Studies 2 and 3), for two reasons. First, there was no reverse-scored item, so the motivation for using the factor scores was reduced (see Footnote 3). Second, because factor scores by definition have a mean of zero, then by standardizing each of our two composites (the GRE bottom-up experience and the high school bottom-up experience), we would eliminate the ability to observe a main effect on this repeated-measures factor. For our between-subjects analyses, we re-conducted them using factor scores and found substantively equivalent results.

⁶ We thank an anonymous reviewer for pointing out that one of our bottom-up items, "think I have learned this before" may not reflect a bottom-up reaction to the task (i.e., "This feels quite familiar to me!") but instead an abstract inference drawn from the manipulation (i.e., "I went to a good high school, so I must have learned this"). Although we intended the item to measure the former, to the extent that some people answered the item consistent with the latter interpretation, then this would merely show an alternate top-down influence on performance estimate (i.e., "Given that my excellent high school offered a comprehensive American history class, I must know this stuff."). But given that participants indicated their feeling that they learned the item before for each individual item, it seems likely that this focused participants on their bottom-up experience with each particular item, instead of on a more abstract inference that would apply to the test as a whole. Nonetheless, the mediation model was significant using each bottom-up item individually, so even if one were to conservatively exclude this particular item, the evidence that bottom-up experience mediates the relationship between self-views and performance estimates does not change.

Between-Subjects Analyses

We should note that the design of Study 4 also allowed us to conduct a between-subjects analysis of mediation, if we set aside the data for the second test that each participant took. Looking at only the first test that participants took, those who took the high school test did not in reality score any higher ($M = 6.5$, $SD = 2.2$) than those taking the GRE test ($M = 7.0$, $SD = 2.1$), $F(1, 156) = 1.75$, $p > .18$. But once again, controlling for actual performance, those taking the high school test perceived that they did better ($M = 5.8$, $SE = 0.29$) than those taking the GRE exam ($M = 4.7$, $SE = 0.30$), $F(1, 155) = 6.55$, $p = .01$. Furthermore, the bottom-up experience of those who took the high school test suggested that they had less difficulty with the test ($M = 2.2$, $SE = 0.060$) than those who completed the GRE test ($M = 2.0$, $SE = 0.060$), $F(1, 152) = 5.60$, $p = .02$. It is worth noting that this between-condition difference is substantially smaller than the one found using the within-subject analysis. This is, of course, not surprising given the within-subjects analysis controls for systematic participant-level error.

To test whether bottom-up experience mediated the impact of test label (high school vs. graduate school) on performance estimates, we regressed performance estimates on test label and bottom-up experience ratings, while controlling for actual performance and which specific version of the test the participant took first. The bottom-up experience strongly predicted performance estimate, $\beta = .66$, $t(151) = 10.55$, $p < .001$, but there was no longer a significant effect of test condition, $\beta = .09$, $t(151) = 1.46$, $p > .14$. A Sobel test confirmed the significance of the full mediation model ($z = 2.31$, $p = .02$).

To provide yet a more stringent test for our hypotheses, we reran our mediation model not simply coding test condition dichotomously but by using the Likert-scale self-view rating provided by each participant. The relevant self-view predicted bottom-up experience, $\beta = .36$, $t(152) = 5.04$, $p < .001$, and performance estimates, $\beta = .38$, $t(152) = 5.42$, $p < .001$. Furthermore, when performance estimate was regressed on both the relevant self-view and the bottom-up experience, bottom-up experience continued to predict performance estimates, $\beta = .61$, $t(151) = 9.32$, $p < .001$, whereas the influence of relevant self-view was reduced, $\beta = .18$, $t(151) = 2.85$, $p = .005$. A much stronger Sobel test confirmed the significance of this partial mediation model ($z = 4.43$, $p < .001$). It is not surprising that this mediation model is much stronger, given that the idiographically measured self-views can account for the influence on the mediator (bottom-up experience) much better than a condition-specific dichotomously coded variable.⁷

In sum, by manipulating within-subjects the self-view relevant to performance evaluation, we found that (a) self-views influenced performance estimates, (b) self-views influenced participants' on-line bottom-up experience, and (c) these differences in on-line bottom-up experiences mediated the link between self-views and performance estimates.⁸

General Discussion

In this article, we set out to examine the role played by top-down self-views and bottom-up experiences in influencing estimates of one's performance. Although past research repeatedly found that chronic self-views drive performance estimates (Ehrlinger & Dun-

ning, 2003), the specific psychological mechanisms responsible for this influence have, to date, remained unclear. In the four studies reported herein, we consistently found that chronic self-views shaped bottom-up experiences with the tests we presented to participants, which then informed their performance estimates. Those who thought they were highly skilled in the domain in question, relative to those who believed they were less skilled, thought they were taking less time to answer questions, were expending less effort, and were feeling more familiarity with the possible answers they could choose.

Specifically, Study 2 showed that the naturally occurring link between chronic self-views and performance estimates were statistically mediated by perceptions of bottom-up experiences. Studies 3 and 4 manipulated which self-view was relevant to a task and showed that bottom-up experiences and performance estimates were influenced by that relevant self-view, but not by an alternative, irrelevant self-view. Again, the impact of relevant self-views on performance estimates importantly involved how participants perceived their bottom-up experiences with the task. Study 3 demonstrated this at the individual participant level; bottom-up experiences were correlated with whether participants rated themselves high or low on the relevant ability. Study 4 demonstrated this at the normative level. When participants were asked about performance on what was supposedly a high school-level history exam, they thought they were more skilled, rated their bottom-up struggles with the exam as more benign, and estimated that they answered more questions correctly.

Finally, Study 1 showed that self-views influence performance estimates if they were manipulated before participants confronted the task, when those self-views could then influence perception of bottom-up experience, but not when they were manipulated after confronting the task, presumably because bottom-up experiences with the task had already been fixed. In addition, Study 3 showed that the influence of these top-down self-views was so strong that, in at least one case, their power far exceeded the power of reality to influence perception. Participants' views of their own ability strongly affected how much time they thought they were taking to

⁷ Given that our between-subjects analytic technique uncovered a between-condition difference in participants' bottom-up experience, does this mean that the mere knowledge that participants would be taking and rating a second test of (expected) different difficulty eliminated the "shifting scale" problem that may have produced the single oddity in the data of Study 3? Possibly, but there is an additional concern that leads us to put more weight in our within-subjects analytic approach. It is possible that those who first took the high school [GRE] test expected the second test to be much more difficult [easy]. This may have pushed the high school and GRE test-takers toward the "I am doing well [poorly]" side of the scales, respectively, leaving themselves room to rate the second test in opposition to the first one. This potential artifact only applies to the between-subjects analyses, which is why we believe the within-subjects analyses are the stronger test.

⁸ We conducted all analyses again using the retrospective reports of bottom-up experiences participants provided at the end of the experimental session. All analyses with these measures led to the same statistical conclusions we reached when focusing on online measures of bottom-up experience. This is not surprising, given that retrospective measures of bottom-up experience were highly correlated with online measures ($r_s = .72$ and $.70$ in the GRE and high school conditions, respectively).

answer questions. Their estimates, however, bore no relationship with how much time they actually took, objectively measured.

The Nature of the Mediation

Careful readers may speculate about an alternative to our mediational model. What if we flipped the roles played by performance estimates and bottom-up cues? Could self-views lead directly to performance estimates, which then inform bottom-up cues? We think this alternative is less plausible than our account given two issues: (a) the order in which bottom-up cues and performance estimates were assessed and (b) the psychology implied by this chronology. First, we asked participants to report their bottom-up experiences as they confronted the test and asked for their performance estimates only after the test was completed. Thus, our measures of bottom-up experience had temporal priority over our measures of performance estimates. That is, bottom-up experiences came first, and it is easier to imagine that they influenced subsequent perceptions of the task (i.e., summary performance estimates) than for those subsequent perceptions to reach back in time and influence reports of bottom-up cues. In any event, work in eyewitness testimony shows that explicitly asking participants to consider their bottom-up experiences, as we did here, inoculates them from top-down influences that may come later in the experience of giving testimony (Wells & Bradfield, 1999).

Had we asked participants to recall their bottom-up experience at the end of the task, this alternative model might have been more psychologically plausible. Given that we measured bottom-up experience online, one would have to posit a hidden variable of a tentative performance estimate that influences ratings of bottom-up experience. That is, for each question, participants first decided their confidence that they had gotten it right and then decided to base their reports of bottom-up experience on this tentative individual-item assessment. Although this process is theoretically possible, it is less plausible than the one we favor, in terms of parsimony.⁹ It is also somewhat hard to imagine that, for example, people decided how much they were concretely vacillating between answer choices only after they decided whether they answered a question correctly, rather than having the experience at the time they worked toward answering the test question.

Thus, across these studies, we found consistent evidence that bottom-up experiences mediated the link between chronic self-views and performance estimates. The individual analyses, however, were inconsistent about whether this mediation was full or only partial. For our theoretical purposes, we think it is essential merely to show partial mediation. Self-views might influence performance evaluations through multiple routes; we wanted to examine whether one of those routes led through the experience of bottom-up cues.

In studies that revealed only partial mediation, we think it is important to keep in mind that our measures of bottom-up experiences may not have been comprehensive. Bottom-up experience is a broad, varied construct, and capturing bottom-up experience as a whole is not a straightforward task. Thus, if we failed to capture all of the bottom-up experiences that were influenced by top-down views, our statistical analyses may have missed the full role being played by bottom-up experience in accounting for the link between self-views and performance estimates. Thus, some mediation may have been partial, not because there was a direct effect of chronic

self-views on performance estimates, but rather because the bottom-up cues mediator was only partially measured. Results from Study 1 buttress this account. By varying when participants learned of the relevant self-view, we could test the extent to which top-down views required subsequent bottom-up experience to contaminate performance estimates, without having to directly measure bottom-up experience. Given that there was not a hint of a difference between the two self-view conditions in the posttest condition, we believe this tilts the scale in favor of the full mediation model.

However, it would be premature to conclude that self-views always exert their effect on performance estimates by altering bottom-up experience. For example, we imagine that there are times in which self-views exert a direct effect on performance estimates because bottom-up cues to task performance are relatively unavailable. An extremely confident stand-up comic performing in front of a live audience may interpret the laughter of her audience as more raucous than it actually is, leading her to conclude she is doing a great job. An extremely confident stand-up comic performing in front of a TV audience will not have as many bottom-up cues to evaluate his performance, and he may decide that he must have done a great job merely by relying upon his prior self-views. We leave it to future research to find additional ways by which self-views may influence performance estimates and in what contexts and for what tasks self-views influence performance estimates in these different ways.

Notes on the Use of Bottom-Up Cues in Performance Estimates

In addition, a critic might also ask whether the apparent heavy reliance on (contaminated) bottom-up experience accurately reflects the way people make judgments in the real world. Unlike in real-world contexts, we impelled participants to explicitly report on their bottom-up experience in a rather heavy-handed way. Could this have led them to artificially attend to and rely on these cues in their performance estimates? We think the answer to this question is “no.” Study 1 provided the strongest evidence that the distorting influence of top-down beliefs on bottom-up cues occurs even in the absence of an experimenter-instructed focus on bottom-up cues. In that study, participants did not report their bottom-up experience, but self-views continued to influence estimates when this information was provided before participants were exposed to the performance task. In contrast, self-views did not influence performance estimates when participants learned of it after the task. This suggests that even in the absence of explicit

⁹ If performance estimates were generated online, one could assume that our measure of performance estimates would be closely tied to online bottom-up experience, rather than to retrospective bottom-up experience measures. If performance estimates were not generated until the end of the test, then performance estimates should be, instead, more closely tied to retrospective bottom-up experience. In a multiple-regression analysis, the retrospective bottom-up experience remains very closely tied to performance estimates, $\beta = .71$, $t(152) = 11.51$, $p < .001$, whereas the online bottom-up experience becomes a much weaker predictor, $\beta = .18$, $t(152) = 2.90$, $p = .004$. This suggests that the performance estimates really were formulated at the end, which is consistent with our favored mediation model.

measures of bottom-up experience, self-views continued to influence performance only when they could influence bottom-up cues to task performance.

A second question is the extent to which self-views override or simply compete for influence with objective reality in shaping bottom-up experience. Consistent with the overriding perspective, Study 3 found that although self-views influenced perceived time perception (i.e., how long it seemed to take to answer a question), actual time passage did not. It seems, however, that some cues may be more constrained by reality than others. Determining the bottom-up cues most likely to be used in different types of performance tasks and the degree to which they are constrained by objective reality will help to determine the circumstances in which self-views are likely to exert their biggest influence on performance evaluation. This issue is particularly important in examining how best to decontaminate one's self-assessments.

A third question is whether bottom-up experience is only a source of error, or whether it is also a source of accuracy. We returned to Studies 2–4 and tested whether any of participants' accuracy (the relationship between actual performance and estimated performance) could be explained through their bottom-up experience. Although actual performance did not relate to bottom-up experience in Study 2, it did in Studies 3 and 4. Indeed, in follow-up analyses, we found that bottom-up experience partially (Study 3) or fully (Study 4) mediated the relationship between actual and estimated performance in the final two studies. But when we regressed bottom-up experience on self-views and actual performance simultaneously, self-views tended to be a stronger predictor. In other words, although bottom-up experience may in part be a function of actual performance, it may equally (if not more so) be a function of misleading self-views. Nonetheless, this point raises hope that we may be able to find cues to self-insight in bottom-up experience, if only future research can determine exactly where we should look.

What conditions must be satisfied for bottom-up experience to become a source of accuracy instead of error? First, the cue must be perceived accurately. The measurement of perceived time (Study 3) that was not related to the actual passage of time shows that accurate detection can be a challenge. But note that even if the perception of a particular bottom-up cue is not distorted by top-down beliefs, two additional conditions must be satisfied for reliance on this cue to increase self-assessment accuracy. Second, the cue must actually be diagnostic of performance. Accurate perception of a nondiagnostic cue is unhelpful. Third, people must have the appropriate naïve theory linking the particular bottom-up cue to actual performance. For example, participants in Studies 2 and 3 believed that solving an item quickly indicated that one had answered it correctly. But when we returned to examine how the actual time it took to solve a problem correlated with performance (in Study 3), we actually found that the longer participants took on a problem, the more likely they were to answer it correctly, $r(131) = .19, p = .03$. Thus, even if participants' assessments of the bottom-up cue (i.e., the passage of time) had been accurate, they would have applied an incorrect theory to understand this cue's actual implications for performance. On an optimistic note, correcting these inaccurate naïve theories may be a relatively simple way to improve accuracy in self-assessment.

In addition, beyond showing that bottom-up experiences mediate the link between self-views and performance estimates, find-

ings across the four studies serve to rule out other ways in which bottom-up cues could influence performance estimates. If, for example, top-down views and bottom-up cues merely exert separate and independent influences on performance estimates, then reliance on top-down cues would not have been sensitive to our timing manipulation in Study 1, nor would the observed mediation have emerged. In addition, if the relevant top-down self-view had its impact through memory, leading people to later recall details of one's performance consistent with one's self-view (cf. Guenther & Alicke, 2008), then, again, the timing manipulation of Study 1 should not have mattered. Furthermore, bottom-up experience assessed online (instead of at the time of performance estimation, when one's memory could be distorted) would not have been a successful mediator.

Implications for Stereotyped Groups

Often, members of stereotyped groups confront a performance-debilitating concern as they perform tasks for which their group is thought to have low ability. Such stereotype threat has been shown to impede African Americans' performance on a test of verbal ability (Steele & Aronson, 1995), the performance of children of low socioeconomic status on an intelligence test (Croizet & Claire, 1998), and White men's performance on a math test that would be compared with Asian men's performance (Aronson et al., 1999). In these performance settings, stereotyped target members are constantly monitoring for cues of how they are performing, assessing whether they are behaving in a stereotype-consistent way (Schmader, Johns, & Forbes, 2008). Some research suggests that this means that they are especially vigilant to actual signs of failure. Neurophysiological techniques like event-related potential and functional magnetic resonance imaging have found that under conditions of stereotype threat, there is evidence of increased monitoring for and vigilance to performance errors (Amodio et al., 2004; Forbes, Schmader, & Allen, 2008).

The four studies contained herein suggest another burden for traditionally stigmatized groups: When a task is ambiguously difficult, negative expectancies tied to stereotypes may lead one to *experience* it as especially difficult, thereby confirming negative stereotypes about oneself and one's group, initiating or reinforcing the negative processes that underlie stereotype threat. Such a subjective sense of failure may only reinforce the anxiety of stereotype confirmation, leading to actually worse performance, thereby exacerbating the negative performance effects of stereotype threat.

In addition, self-views' coloring of bottom-up experience may have other negative longer term impacts on stereotype-relevant tasks. Ehrlinger and Dunning (2003) described how the link between self-views and performance assessments may underlie differences in the enthusiasm with which men and women pursue scientific careers. Replicating past work, Ehrlinger and Dunning found that women thought less of their scientific talent than did men and that this difference led to women thinking they had done more poorly on a pop quiz on science than did men, even though there were no differences in objective performance. This perception of performance, however, mattered in that it led women to volunteer less often than did men for a science game show taking place later on. In short, lowered estimates of performance led to greater lack of interest in science.

Given our data, we can speculate on an additional route by which preconceived self-views can prompt disinterest in a relevant activity. Even if people with negative self-views are provided with objective feedback about their performance, indicating they had done well, their negative self-views may still leave them with the impression that the process followed to achieve that performance was aversive and laborious. That is, someone might be told that she did well on a science quiz, but she may still not want to go on in science because self-views lead her to believe that she struggled with the quiz. To be sure, we did not explore this possibility in the research described in this article, but it would be interesting to pursue whether perceptions of bottom-up cues shape not only performance evaluations but also subsequent interest in the activity itself.

Conclusion

The top-down influence of chronic self-views does not vie with bottom-up experience in influencing self-evaluation. Rather, self-views define bottom-up experiences. Even though self-estimates of task performance closely aligned with self-reported bottom-up experience with the task, this experience was constructed, at least in part, on the basis of preconceived self-notions coming into the task. This research may seem to paint a somewhat bleak picture for the ability of people to evaluate themselves accurately, but we hope that future research can uncover what objective cues actually are most diagnostic of task performance. Until then, it remains important that our work be evaluated by those who do not have clear top-down beliefs about what to expect from us, even if they at times don't realize just how great that work is.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Allison, R. I., & Uhl, K. P. (1964). Influence of beer brand identification on taste perception. *Journal of Marketing Research*, *1*, 36–39.
- Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartely, S. L., & Covert, A. E. (2004). Neural signals for the detection of intentional race bias. *Psychological Science*, *15*, 88–93.
- Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology*, *27*, 576–605.
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, *35*, 29–46.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.
- Belmore, S. M. (1987). Determinants of attention during impression-formation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 480–489.
- Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition: The 27th Carnegie Symposium on Cognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.
- Biernat, M. (2005). *Standards and expectancies: Contrast and assimilation in judgments*. New York: Psychology Press/Taylor & Francis.
- Bunz, U., Curry, C., & Voon, W. (2007). Perceived versus actual computer-email-web fluency. *Computers in Human Behavior*, *23*, 2321–2344.
- Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, *13*, 225–245.
- Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 142–150.
- Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype and threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, *24*, 588–594.
- Crosby, R. A., & Yarber, W. L. (2001). Perceived versus actual knowledge about correct condom use among U.S. adolescents: Results from a national study. *Journal of Adolescent Health*, *19*, 134–139.
- Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press.
- Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*, 69–106.
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, *84*, 5–17.
- Fiske, S. T., Neuberg, S. L., Beattie, A. E., & Milberg, S. J. (1987). Category-based and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. *Journal of Experimental Social Psychology*, *23*, 399–427.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw Hill.
- Forbes, C. E., Schmader, T., & Allen, J. J. B. (2008). The role of devaluing and discounting in performance monitoring: A neurophysiological study of minorities under threat. *Social Cognitive and Affective Neuroscience*, *3*, 253–261.
- Guenther, C. L., & Alicke, M. D. (2008). Self-enhancement and belief perseverance. *Journal of Experimental Social Psychology*, *44*, 706–712.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self supervisor, self peer, and peer supervisor ratings. *Personnel Psychology*, *41*, 43–62.
- Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods*, *6*, 115–134.
- Jussim, L., Coleman, L., & Nassau, S. (1987). The influence of self-esteem on perceptions of performance and feedback. *Social Psychology Quarterly*, *50*, 95–99.
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, *32*, 1–24.
- Lee, L., Frederick, S., & Ariely, D. (2006). Try it, you'll like it: The influence of expectation, consumption, and revelation on preferences for beer. *Psychological Science*, *17*, 1054–1058.
- Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, *135*, 501–512.
- Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, *15*, 374–378.
- Lindeman, M., Sundvik, L., & Rouhiainen, P. (1995). Under- or overestimation of self? Person variables and self-assessment accuracy in work settings. *Journal of Social Behavior and Personality*, *10*, 123–134.
- Mabe, P. A., III, & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, *67*, 280–296.
- Makens, J. C. (1965). Effect of brand preference upon consumers' perceived taste of turkey meat. *Journal of Applied Psychology*, *49*, 261–263.

- Marteau, T. M., Johnston, M., Wynne, G., & Evans, T. R. (1989). Cognitive factors in the explanation of the mismatch between confidence and competence in performing basic life support. *Psychology and Health, 3*, 173–182.
- Mast, M. S. (2005). Interpersonal hierarchy expectation: Introduction of a new construct. *Journal of Personality Assessment, 84*, 287–295.
- McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron, 44*, 379–387.
- McFarland, C., Cheam, A., & Buehler, R. (2007). The perseverance effect in the debriefing paradigm: Replication and extension. *Journal of Experimental Social Psychology, 43*, 233–240.
- Nevid, J. S. (1981). Effects of brand labeling on ratings of product quality. *Perceptual and Motor Skills, 53*, 407–410.
- Olson, J. C., & Dover, P. A. (1978). Cognitive effects of deceptive advertising. *Journal of Marketing Research, 15*, 29–38.
- Peng, K., Nisbett, R., & Wong, N. (1997). Validity problems comparing value across cultures and possible solutions. *Psychological Methods, 2*, 329–344.
- Plassman, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences, USA, 105*, 1050–1054.
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*, 338–342.
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perceptions and social perception: Biased attributional processing in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*, 880–892.
- Sagar, H., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology, 39*, 590–598.
- Sanford, A. J., Fay, N., Stewart, A., & Moxey, L. (2002). Perspective in statements of quantity, with implications for consumer psychology. *Psychological Science, 13*, 130–134.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*, 336–356.
- Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1074–1083.
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14*, 332–348.
- Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 39, pp. 127–161). San Diego, CA: Elsevier.
- Shrauger, J. S., & Terbovic, M. L. (1976). Self-evaluation and assessments of performance by self and others. *Journal of Consulting and Clinical Psychology, 44*, 564–572.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.
- Tracey, J. M., Arroll, B., Richmond, D. E., & Barham, P. M. (1997). The validity of general practitioners' self assessment of knowledge. *British Medical Journal, 315*, 1426–1428.
- von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1995). On the role of encoding processes in stereotype maintenance. *Advances in Experimental Social Psychology, 27*, 177–254.
- Wansink, B., Park, S. B., Sonka, S., & Morganosky, M. (2000). How soy labeling influences preference and taste. *International Food and Agribusiness Management Review, 3*, 85–94.
- Wardle, J., & Solomons, W. (1994). Naughty but nice: A laboratory study of health information and food preferences in a community sample. *Health Psychology, 13*, 180–183.
- Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitness recollections: Can the postidentification-feedback effect be moderated? *Psychological Science, 10*, 138–144.

Received July 30, 2008
 Revision received July 3, 2009
 Accepted July 7, 2009 ■

E-Mail Notification of Your Latest Issue Online!

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