Lessons From the Past: Do People Learn From Experience That Emotional Reactions Are Short-Lived?

Timothy D. Wilson
Jay Meyers
University of Virginia
Daniel T. Gilbert
Harvard University

Do people learn from experience that emotional reactions to events are often short-lived? Two studies indicate that it depends on whether the events are positive or negative. People who received positive or negative feedback on a test were not as happy or unhappy as they would have predicted. People in the positive feedback condition did not learn from this experience when making predictions about their reactions to future positive events. People in the negative feedback condition moderated their predictions about their reactions to future negative events, but this may not have been a result of learning. Rather, participants denigrated the test as a way of making themselves feel better and, when predicting future reactions, brought to mind this reconstrual of the test and inferred that doing poorly on it again would not make them very unhappy. Experience with a negative event (but not with a positive event) may improve the accuracy of one's affective forecasts, but the extent to which people learn from their affective forecasting errors may be limited.

"I'll be elated for days if we beat the Rams," a sports fan thinks, even though he or she should know better. Sporting events typically do not have long-lasting effects on people's happiness. Wilson, Wheatley, Meyers, Gilbert, and Axsom (2000), for example, assessed college football fans' happiness 1 to 3 days after their favorite teams won or lost a game and found that the outcomes of the games had no detectable effect on their happiness; yet, the fans predicted that the outcome of the games would influence their happiness for several days.

This *durability bias*—the tendency to overestimate the duration of one's emotional reactions to future events—has been found in a variety of settings with a variety of populations (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998; Wilson et al., 2000). Curiously, it occurs in domains in which people have considerable past experience. Wilson et al.'s (2000) participants had watched many football games, and yet they overpredicted the impact of future ones on their happiness. Similarly, people expect new television sets, upgraded computers, and fancy cars to cause lasting pleasure when their previous wide-screen televisions, power notebooks, and luxury sedans did not. The purpose of the present studies was to investigate why people fail to generalize from their past experiences when making affective forecasts and the conditions under which they will generalize, at least to some extent.[1]

On the face of it, it seems easy to learn from past emotional experiences. People simply need to recall how they felt after a similar event in the past, such as the fact that the thrill of their new 27-inch television lasted only a couple of weeks, and apply that knowledge to the future, predicting that the thrill of owning a new 32-inch model will probably wear off quickly as well. We believe, however, that generalization of this sort can be quite difficult.

Authors' Note: We gratefully acknowledge the support of the National Institute of Mental Health (Research Grant No. RO1-MH56075) to Daniel T. Gilbert and Timothy D. Wilson. Correspondence concerning this article may be sent to Timothy D. Wilson, Department of Psychology, University of Virginia, P.O. Box 400400, Charlottesville, VA 22904-4400; e-mail: tdw@virginia.edu.

PSPB, Vol. 27 No. 12, December 2001 1648-1661

Necessary Conditions to Learn From Past Emotional Experiences

To learn from their past emotional experiences, people would have to satisfy three conditions. First, they would need to make an effort to compare their past experiences with future ones, rather than simply thinking about a future event in isolation with no reference to the past (mental effort criterion). Buehler and McFarland (2001), for example, found that when people were induced to focus on a future event without thinking about similar past events, they made less accurate affective forecasts. Second, if people do make the effort to consult the past, they need to decide which of their many past experiences are most applicable to the future event they are considering (applicability criterion). Third, even if people recognize which events are applicable and make an effort to think about them, they need to recall or reconstruct accurately how they felt after those events (accuracy criterion).

We suggest that people often fail to satisfy one or more of these conditions. People often opt for simple mental strategies over effortful ones, suggesting that in many situations, they might fail to satisfy the mental effort criterion (e.g., Gilbert, 1991; Petty & Cacioppo, 1986). If people do think about the past, they might not recognize which events are most relevant to future ones, failing to satisfy the applicability criterion. Suppose that football fans are wondering how long they will be happy after their favorite team beats the Rams next month. Is the best predictor how happy they were after their team beat the Rams 2 years ago, how happy they were after their team beat the Lions last week, or how happy they were after their bowling team beat the Elks Lodge last night? No two events are identical and it is not always obvious which past events are most relevant to one's future reactions (Higgins, 1996; Koehler, 1996).

Even if people decide that a past event is applicable, they might not recall how happy they were after this event, failing to satisfy the accuracy criterion. There is evidence that people's memory for emotional reactions is poor. Christianson and Safer (1996) went so far as to conclude that "There are apparently no published studies in which a group of subjects has accurately recalled the intensity and/or frequency of their previously recorded emotions" (p. 235). People have been shown to have poor memory for the intensity or frequency of specific emotions such as anger and sadness (Levine, 1997), pain (Erskine, Morley, & Pierce, 1990), positive and negative moods (Thomas & Diener, 1990), and attitudes (Ross, 1989; Ross & Newby-Clark, 1998).

It thus seems unlikely that people will typically satisfy all three conditions necessary to learn from their past emotional experiences. There is reason to believe, however, that people learn different lessons from positive and negative events. We argue that people often do not learn from positive events because they fail to meet one or more of the criteria for learning to occur. In contrast, people sometimes appear to learn from negative events, not necessarily because they meet the necessary criteria but because negative experiences change their construal of the event in ways that moderate their predictions for the future.

Failing to Learn From Positive Events

One reason people fail to generalize from positive events is that as time goes by, they fail to remember that these reactions were short-lived. Meyers, Wilson, and Gilbert (2000) and Mitchell, Thompson, Peterson, and Cronk (1997) found evidence for a retrospective durability bias, whereby people overestimated how happy they were after positive emotional events in the past. In one study, people interested in politics were surveyed at three points in time: a few weeks before the 1996 presidential election, right after the election, and 3 months later. Democrats showed a strong durability bias before the election, predicting that they would be much happier following Bill Clinton's victory than they in fact were. They also showed a retrospective durability bias, whereby 3 months after the election they recalled being happier following Bill Clinton's victory than they in fact had been. Similarly, participants in studies by Mitchell et al. (1997) predicted that enjoyable activities such as a bicycle trip would be significantly more enjoyable than they in fact were, and they remembered that the activities were more enjoyable than they in fact were.

Meyers et al. (2000) suggested that there is a common mechanism responsible for both the retrospective and prospective durability bias, namely, *focalism*, whereby people think too much about the event in question and fail to consider the consequences of other events that are likely to occur. When predicting how they will feel in the future after an emotional event, people think too much about that event and too little about other events that will influence their thoughts and feelings (Wilson et al., 2000). Similarly, when people try to reconstruct how happy they were in the past, they appear to focus too much on the event and not enough on other events that occurred at the time, thereby overestimating the emotional impact of the event. This retrospective durability bias helps to explain why people do not learn from their past experiences when predicting the future. One of the mechanisms that produces the durability bias (focalism) appears to operate in prospect and retrospect.[2]

Sometimes, of course, people think about the future right after experiencing an emotional event, when their reaction to it is fresh in their minds. Under these conditions, the accuracy criterion is likely to be satisfied; people can easily recall that their emotional reaction was rel-

atively short-lived ("Hmm, the new 27-inch television has not made me happy for as long as I expected") and thus might generalize from this fact when forecasting their future reactions ("I guess a new 32-inch model wouldn't make me happy for very long either").

Even if they can recall how they felt in the past, however, people still might fail to satisfy the mental effort criterion; that is, they might not focus on their past feelings and apply this knowledge to their predictions about the future. There is evidence that when people experience neutral or positive events, they are not very motivated to analyze the causes of their feelings (Hastie, 1984; Wong & Weiner, 1981). Generalizing from these findings, we hypothesized that when people experience a neutral or positive event and then are asked to predict how they will feel after similar events in the future, they will not engage in an effortful recall and application of their past reactions. Instead, they will focus on the future event and assume that it will have a long-lasting effect ("Wow, a new 32-inch TV, that would be great").

To test these hypotheses, we randomly assigned people to receive positive or no feedback on a test of social aptitude and then asked them to predict how happy they would be if they did well on similar future tests. Consistent with prior research on the durability bias, we expected that this feedback would not influence people as much as they expected. Because people made forecasts about future events soon after experiencing similar ones, they should be able to recall accurately that these events did not influence them very much. Even so, we hypothesized that people who received positive feedback would not generalize from this experience and would predict that similar positive events in the future would make them quite happy.

We do not mean to suggest that people will never recall and apply how they felt after past positive events. If the event that was experienced or the one to be predicted is extremely consequential or unexpected, people might well go to the effort of recalling and applying their past reactions. In the present studies, however, people were not expected to experience extremely discrepant reactions, and the events they predicted were not extremely consequential. Under these conditions, we suggest, people might not engage in the necessary mental effort to learn from their recent positive experiences.

Reconstruing Negative Events

In contrast, we predicted that after experiencing a negative event, such as getting a poor grade on a test, people would moderate their forecasts about how similar events would influence them in the future. One reason for such generalization might be that people genuinely learn from the fact that they did not have as negative a reaction as they expected and apply this knowledge to the future. Negative experiences often have less effect on people's happiness than they expect; in fact, we have found that this absence of a negative effect is more unexpected to people than the absence of a positive effect (Meyers et al., 2000). Consequently, people might be more likely to think about why they did not feel as negatively as they expected, leading them to meet the necessary criteria to learn from experience: They think about their reaction to doing poorly on a test (mental effort); they recall accurately that they did not feel very upset (accuracy); and they apply this knowledge to the future, predicting that they will not feel upset if they do poorly on similar tests in the future (applicability).

Although people may well go through these steps under some circumstances, we suggest that there is a simpler, less effortful way that receiving negative feedback can moderate people's predictions about their future reactions. Such generalization could occur not because people satisfy the conditions necessary for genuine learning but because the negative experience triggers a reconstrual of the event. It is well known that people possess a powerful "psychological immune system" that speeds recovery from negative experiences (e.g., Festinger, 1957; Taylor, 1991). People rationalize and reconstrue negative events in ways that dampen the pain they cause. For rationalization to be most effective, it is helpful that people not be aware that they are doing it (Gilbert et al., 1998). One consequence of such immune neglect—the tendency to be unaware of the operation of one's own psychological immune system—is the durability bias. Because people do not recognize how much they will transform negative events psychologically in ways that ameliorate negative feelings, they overestimate how long it will take them to recover from the events (Gilbert et al., 1998).

We suggest that after a negative event has occurred, people's reconstrual of it also will influence their forecasts about how they will feel after similar, future events. That is, instead of consulting and learning from their prior reactions, people simply might bring to mind their new, benign construal of the experience and predict that such a mild event would not influence them very much. In other words, people might moderate their predictions not because they meet the criteria necessary to learn from experience but because their view of the test has changed.

To test this reconstrual hypothesis, we randomly assigned some participants to receive negative feedback on the test of social aptitude and asked them to predict how happy they would be if they did poorly on similar tests in the future. We hypothesized that people would rationalize their poor performance such that it did not make them as unhappy as they would have predicted

and, further, that this rationalization would cause them to predict that a future negative performance on the test would not make them very unhappy either.

We should note that the learning-from-experience and reconstrual hypotheses are not mutually exclusive. People might predict that a future negative performance will have little impact because (a) they remember that they did not feel badly after doing poorly the first time and apply this knowledge to the future and (b) their view of the test has changed, leading them to predict that doing poorly on an invalid test will not affect them very much. The goal of the present studies was to test the latter reconstrual view to see if this relatively simple process can account for changes in people's predictions about future negative events. We will return to the learning-by-experience hypothesis in the General Discussion and discuss evidence relevant to it.

In addition to asking people how happy they would be after experiencing similar events in the future, we asked them to predict how they would feel if they received positive or negative feedback in situations that were different from the test they had just taken (e.g., overhearing a neighbor complain that they were insensitive to other people). We predicted that the extent to which people who received negative feedback would generalize to these different situations would depend on the specificity of their rationalizations about their poor performance on the test. If people limit their rationalizations to devaluing the specific test, then they should generalize only to similar situations (e.g., taking similar tests in the future). If they make more general rationalizations, such as devaluing the test and deciding that social aptitude is not a very important skill, then they should generalize to both the similar and dissimilar future negative events (e.g., the neighbor's negative comment), predicting that all of them will have less impact than they will for those people who did not receive negative feedback.

STUDY 1: GENERALIZING FROM POSITIVE AND NEGATIVE EXPERIENCES

Method

OVERVIEW

People received positive, negative, or no feedback on a test of social aptitude and then predicted how happy they would be after receiving negative or positive feedback in a variety of future situations. Some of these situations were very similar to taking the social aptitude test, whereas others were quite different. The study thus employed a 3 (feedback: positive, negative, none) × 2 (similarity of future situations: similar vs. dissimilar) × 2 (valence of future events, positive vs. negative) design, with repeated measures on the last two factors. We hypothesized that people in the negative feedback condition would predict that they would be less affected by future negative events than people in the positive and no feedback conditions predicted they would be. No differences between conditions were expected for future positive events; people in the positive feedback condition should say that they would be as happy after these events as people in the negative and no feedback conditions.

PARTICIPANTS

The participants were 128 undergraduates enrolled in psychology courses at the University of Virginia who received course credit for their participation. Before participating, people completed the Beck Depression Inventory (BDI) and a baseline measure of happiness. To avoid the possibility that negative feedback would be distressing to people who were depressed, people who rated their current happiness as 3 or less or scored 11 or more on the BDI were not included in the study. We ran these people in the no feedback condition so as not to single them out but did not include their data in the analyses. Twenty-two participants met these criteria. The final sample thus consisted of 106 participants (88 women and 18 men).

PROCEDURE

Individual participants were asked to take a test called the Virginia Social Aptitude Scale (V-SAS), which was ostensibly a newly developed test of "social aptitude . . . the ability to perceive and interpret the social behavior of others." We deliberately made the description of this trait somewhat ambiguous to make it easy for people who did poorly on the test to decide that it was not a very important skill. The experimenter mentioned that because many students had expressed an interest in knowing how well they had done on the test, they would receive feedback in the form of a letter grade. They were given a sheet that ascribed the following descriptors and percentiles to grades A through D: "A" or "Outstanding," 90th-100th percentile; "B" or "Good," 70th-89th percentile; "C" or "Fair," 50th-69th percentile; "D" or "Below Average," 30th-49th percentile.

Participants rated their current level of happiness ("In general, how happy would you say you are today, compared to how happy you are ON AVERAGE?" 1 = *below average happiness*, 5 = *average happiness*, 9 = *above average happiness*) and some filler personality scales (e.g., the Rosenberg Self-Esteem Scale). They then took the V-SAS, which involved looking at 40 black-and-white photocopied pictures of faces and guessing which emotion each person was expressing from a list of eight emotions and emotion blends. The faces, which were from Ekman and Friesen (1975), were pretested on an independent group of students who were asked to identify the emotion without the benefit of multiple-choice options. Based on the pilot results, we constructed options for

each face that were difficult. After 10 minutes, the experimenter collected the test and scored it in a different room.

Feedback manipulation. In the negative feedback condition, the experimenter scored the test by marking 19 of the 40 items incorrect and assigned it a grade of D. Based on pilot testing, the items that generated the most inconsistent responses were marked wrong. In the positive feedback condition, the experimenter marked 5 of the 40 items incorrect and assigned it a grade of A. In both conditions, she gave the participant the graded answer sheet and the description of the percentiles for each grade in a folder and said that she had to go make a photocopy of the next questionnaire. Participants in the no feedback condition were told that they would receive their grade at the end of the study.

DEPENDENT MEASURES

Happiness measures. Participants rated their happiness 5 minutes after receiving their feedback on the same scale as they had rated their baseline happiness. Most participants then filled out an indirect measure of their mood—the Associated Reasoning Scale (Mayer & Hanson, 1995). (We added this scale after the study had begun; 74 of the 106 participants completed it.) This scale asks people to rate the probability that positive and negative events will occur, such as an improvement in the economy or the divorce of a married couple, and to choose a sample member of various categories (e.g., "a type of worker") from a list of positive and negative examples (e.g., conscientious vs. lazy). Previous studies have found that responses to this scale correlate reasonably well with self-reported mood and are sensitive to experimental manipulations of mood (e.g., Mayer & Hanson, 1995).

Before making their predictions about how happy they would be in future situations, it was important that participants' current mood be equivalent in each condition (to avoid the possibility that their current moods would color their predictions about the future). We asked all participants to read articles that were intended to neutralize any lingering effects of the feedback. They were told that future versions of the V-SAS might include examining people's reactions to written stimuli and were asked to read two articles and answer questions related to them. The first article was a consumer report about compact disc players, whereas the second article was a piece by Dave Barry called "Consumers From Mars" that discussed advertising and marketing in a humorous manner. After reading each article, participants answered filler questions about how well written, enjoyable, interesting, informative, and entertaining the article was and the extent to which it kept their attention. All participants then received a questionnaire on which they again indicated their current happiness on the same happiness scale as before. This question was answered approximately 15 minutes after taking the V-SAS test. Next, participants filled out a second form of the Associated Reasoning Scale. The forms of this scale that people completed first and second were counterbalanced.

Prediction questions. Participants then received the main dependent measures, on which they predicted how happy they would be after receiving positive and negative feedback on a variety of tests and in other situations. They did so on the same 9-point scales on which they had rated their actual happiness earlier, where 1 = *below average happiness*, 5 = *average happiness*, and 9 = *above average happiness.* Participants were told to imagine the events occurred in 3 weeks and to judge each separately and independently of the others. They also were told that the abilities measured on the tests were not related to one another and that performance on one test was not predictive of performance on the other tests.

There were five similar events that involved taking another version of the V-SAS. The first test was described as identical to the one people had just taken except that it used a different set of black and white photographs. The next four were all subtests of the V-SAS that varied on at least one dimension from the subtest people had taken: one used high-quality color photos instead of black-and-white photos, one involved judging people's personalities (instead of their emotions) from a set of color photos, one involved judging people's emotions from recordings of their tone of voice instead of photographs, and one involved judging people's personality from recordings of their tone of voice. All five were clearly labeled as subtests of the V-SAS. People predicted how happy they would be 5 minutes after receiving an A and a D on each test.

The four dissimilar events involved receiving positive or negative feedback in situations other than taking the V-SAS: overhearing a stranger in a restaurant comment how "sensitive to others" you are (positive) or overhearing the stranger complain how "insensitive of others you are" (negative); being told by a teaching assistant that your work shows great creativity (positive) or that your work lacks creativity (negative); hearing that your neighbor told your landlord that you are "kind and sensitive to everyone around you" or that your neighbor complained that you were "loud and insensitive to those around you"; and getting an A on a Shape Manipulation Test of Creativity (in which the task is to reconfigure shapes to solve various problems) or getting a D on the Shape Manipulation Test of Creativity. In each case, people rated how happy they would be 5 minutes after the positive and negative version of the event. Note that each of these events was similar in at least one way to the V-SAS; for example, some involved feedback on how sensi-

tive people were (the same trait measured by the V-SAS), whereas others involved feedback on tests of other abilities or in other academic situations. However, unlike the similar events, none involved taking a subtest of the V-SAS.

Rationalization measure. Participants then answered several questions designed to assess how much they had rationalized a poor performance on the V-SAS, including how valid the test was (1 = *not very valid*, 9 = *extremely valid*), its fairness (1 = *not very fair*, 9 = *extremely fair*), how easy it was to concentrate while taking the test (1 = *very difficult*, 9 = *very easy*), how easy it is to measure social aptitude (1 = *very difficult*, 9 = *very easy*), how important psychology experiments are (1 = *not very important*, 9 = *extremely important*), how much something unusual (such as how tired they were) influenced their test performance (1 = *not at all*, 9 = *very much*), their opinions of multiple-choice tests (1 = *very poor test format*, 9 = *very good test format*), and their physical health (1 = *not so great*, 9 = *just fine*). They also rated the importance of the following outcomes and abilities (all using a scale that ranged from 1 = *not very important* to 9 = *extremely important*): the importance of doing well on the face-reading test, face-reading ability, being a good face reader in the work place, being a good face reader when dealing with people, and being a good face reader in friendships. All participants were then fully debriefed.

Results and Discussion

People's initial level of happiness in the negative feedback, no feedback, and positive feedback conditions were 5.38, 6.00, and 5.82 (*SD*s = 1.06, 1.17, 1.24), respectively, $F(2, 103) = 2.79$, $p = .07$. The mean in the positive feedback condition did not differ significantly from the mean in the negative feedback condition, $F(1, 103) = 2.55$, $p = .11$. To control for individual differences in initial happiness, we subtracted people's baseline happiness reports from their happiness reports after taking the test.

EFFECTS OF FEEDBACK ON ACTUAL HAPPINESS

People reported their happiness 5 and 15 minutes after taking the test. At 5 minutes, there was a tendency for people in the positive feedback condition to be happier than people in the no feedback and negative feedback conditions (*M*s = .09, –.17, and –.19, respectively; *SD*s = .38, .70, .62, respectively). However, the main effect of feedback did not reach significance, $F(2, 103) = 2.41$, $p = .10$. At 15 minutes, the means were .12, –.22, and .00, respectively (*SD*s = .70, .72, .67, respectively), and again, the main effect of feedback was not significant, $F(2, 103) = 2.19$, $p = .12$. A 3 (feedback: negative, none, positive) × 2 (time: 5 vs. 15 minutes) mixed-model ANOVA did not reveal any significant effects: main effect of feedback, $F(2, 103) = 2.29$, $p = .11$; Feedback × Time interaction, $F(2, 103) = 2.19$, $p = .12$.[3]

People's responses on the Associated Reasoning Scale (Mayer & Hanson, 1995) also indicated that the feedback had little, if any, effect on people's mood. There was neither a main effect of feedback nor a Feedback × Time interaction on this measure, *F*s(2, 71) ≤ 1. There was a nearly significant effect of time, $F(1, 71) = 3.39$, $p = .07$, reflecting the fact that people's judgments were more positive at 5 than 15 minutes (*M*s = 23.16 and 21.39, *SD*s = 6.89 and 7.15). The absence of a feedback effect on the Associated Reasoning Scale is, of course, a null effect that should be interpreted cautiously, especially because there was a reduced sample size on this measure. It is worth noting, however, that other studies have found that this scale correlates reasonably well with self-reported mood and is sensitive to experimental manipulations of mood (e.g., Mayer & Hanson, 1995). In our study, the scales correlated significantly (albeit modestly) with self-reported happiness averaged across 5 and 15 minutes, $r(72) = .26$, $p = .02$.

THE PREDICTION DURABILITY BIAS

We could test for a prospective durability bias in the no feedback condition because these participants predicted how they would feel about getting an A or a D on the same test they had taken that day (with new photographs) without having received any prior feedback. As hypothesized, they predicted that they would be relatively happy if they got an A ($M = .78$, $SD = 1.31$), which was significantly higher than positive experiencers reported feeling after actually getting an A ($M = .09$, $SD = .38$), $t(67) = 2.89$, $p = .005$. Similarly, they predicted that they would be unhappy if they received a D ($M = –2.31$, $SD = 1.62$), which was significantly lower than negative experiencers reported feeling after actually getting a D ($M = –.19$, $SD = .62$), $t(71) = 7.42$, $p < .001$. No feedback participants, of course, were making predictions for a somewhat different situation (taking the test in the future for the second time) than people in the positive and negative feedback condition actually experienced. It is thus worth noting that the magnitude of their apparent durability bias was nearly identical to that found by Meyers et al. (2000) among forecasters who made predictions for the same situation that people in the positive and negative feedback conditions experienced. These results suggest that people's emotional reactions to doing well or poorly on the test did not last as long as they would have predicted.

RATIONALIZATION

An iterative principal axis factor analysis of the 13 items on the rationalization questionnaire, with a varimax rotation, yielded a sensible three-factor solution (with eigenvalues > 1). One factor consisted of four

items about the specific test people had taken that day (e.g., "How valid do you think the face-reading test of empathy is?), a second factor consisted of four items about the general importance of social aptitude in different settings (e.g., "How important is being a good face reader when dealing with people?"), and a third factor consisted of three items about people's physical and psychological state when they took the test (e.g., "How easy was it for you to concentrate while taking the test?"). Two items (how easy it is to measure social aptitude and their opinions of multiple-choice tests) did not load strongly on any of the three factors and were dropped from the analyses.[4] People's responses were reverse scored such that high numbers revealed more rationalization (e.g., more negative ratings of the validity of the test).

A series of ANOVAs revealed that negative experiencers had higher scores on each of the rationalization indices. They rated the specific test more negatively (M = 5.43, SD = 1.11) than people in the no feedback and positive feedback conditions (Ms = 4.27 and 4.42, SDs = 1.32 and 1.38), $F(2, 103) = 8.98$, $p < .001$. They said that social aptitude was a less important skill than did people in the no feedback and positive feedback conditions (Ms = 2.78 vs. 2.26 and 2.20; SDs = 1.32, 1.10, .93, respectively), $F(2, 103) = 2.83$, $p = .06$, and they also said that they were in a worse psychological and physical state while taking the test than did people in the no feedback and positive feedback conditions (Ms = 4.01 vs. 3.61 and 2.93; SDs = 1.87, 1.64, 1.58, respectively), $F(2, 103) = 3.55$, $p = .03$. These results suggest that negative experiencers rationalized their poor performance broadly, derogating the specific test they had taken, downplaying the importance of social aptitude as a trait, and claiming that they were not in an optimal physical or psychological state while taking the test.

PREDICTIONS ABOUT REACTIONS TO FUTURE POSITIVE AND NEGATIVE EVENTS

People's predictions for the five similar future events (i.e., the five V-SAS subtests) were highly correlated (α = .93), as were their predictions for the four dissimilar similar events, (α = .90). Therefore, we averaged their ratings within each category. As seen in Figure 1, our hypotheses were largely confirmed. First, as hypothesized, people who received positive feedback did not generalize at all to the future experiences. Their affective forecasts were very similar to the forecasts made by no feedback participants; a 2 (actual feedback: positive vs. no feedback) × 2 (anticipated future feedback: positive vs. negative) × 2 (situations: similar vs. dissimilar) ANOVA did not reveal any significant main effects or interactions involving actual feedback, Fs(1, 67) < 1. The only significant effects were an unsurprising main effect of anticipated future feedback, $F(1, 67) = 533.45$, $p < .001$, reflecting the fact that people predicted they would be happier if they received positive feedback in the future situations than if they received negative feedback, and an Anticipated Future Feedback × Situations interaction, $F(1, 67) = 129.20$, $p < .001$, reflecting the fact that people predicted that the similar test situations would have less impact (e.g., less of a difference between receiving positive vs. negative feedback) than the dissimilar situations would (see Figure 1).

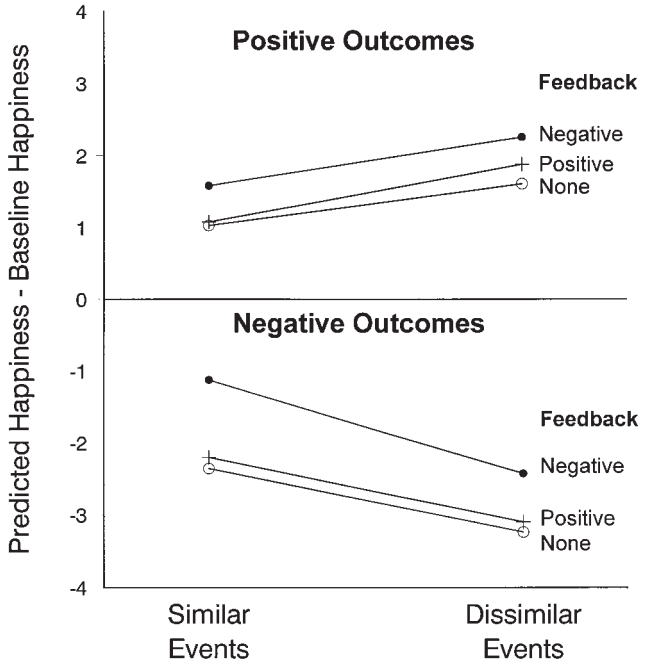

Figure 1 Study 1: Predicted happiness following similar and dissimilar future events.

In short, as hypothesized, receiving positive feedback on the V-SAS test had no detectable effect on people's predictions about how happy they would be in similar or dissimilar situations in the future. Another way of illustrating this is to examine people's predictions about how they would feel if they got an A on the same test that they had just taken (this was the first situation they were asked to predict). Even though they had reported that they were only .09 (SD = .38) points above baseline 5 minutes after receiving an A on the test, they predicted that they would be .97 (SD = 1.45) points higher than their baseline level of happiness 5 minutes after getting an A on the same test in the future, $t(32) = 3.90$, $p < .001$.

Also as hypothesized, negative experiencers predicted that future negative feedback would not make them as unhappy as people in the no feedback condition said it would. A 2 (actual feedback: received a D vs. no feedback) × 2 (anticipated future feedback: positive vs. negative) × 2 (situations: similar vs. dissimilar) ANOVA revealed a nearly significant three-way interaction, $F(1,$

71) = 3.33, $p = .07$, which we decomposed by conducting separate 2 (actual feedback: received a D vs. no feedback) × 2 (situations: similar vs. not similar) ANOVAs for predicted reactions to future positive versus negative events. For negative future events, there was a highly significant effect of actual feedback, $F(1, 71) = 10.66$, $p = .002$, reflecting the fact that people in the no feedback condition predicted that they would feel unhappier after future negative feedback than did negative experiencers (Ms = –2.80 vs. –1.78; SDs = 1.47, 1.19). As expected, negative experiencers moderated their forecasts more for similar than dissimilar situations, as reflected by a significant Actual Feedback × Situations interaction, $F(1, 71) = 3.90$, $p = .05$. However, the difference between the negative feedback and feedback conditions was significant for both the similar and dissimilar situations, Fs(1, 71) = 13.47 and 6.20, ps < .001 and .03, respectively. Thus, whereas negative experiencers moderated their forecasts more for similar than dissimilar future events, this moderation was significant in both cases.

As seen in Figure 1, negative experiencers did not generalize completely from their lack of reaction to doing poorly on the test. Their predicted happiness 5 minutes after getting a D on the same test in the future was –.84 ($SD = 1.26$), which was significantly lower than how happy they reported being 5 minutes after getting a D, –.19 ($SD = .62$), $t(36) = 3.82$, $p < .005$. Nonetheless, their predictions about how happy they would be after future negative events were more moderate than the predictions in the no feedback and positive feedback conditions, indicating that some generalization occurred.

For positive future events, people who received negative feedback made more positive predictions than did no feedback people; main effect of actual feedback, $F(1, 71) = 4.63$, $p = .04$. This finding is similar to a contrast effect that Meyers et al. (2000) found, whereby negative experiencers said they would have felt better if they had received an A on the test than positive experiencers reported feeling. After a negative experience, positive ones seem even sweeter. The Actual Feedback × Situations interaction was not significant, $F(1, 71) < 1$.

MEDIATION ANALYSES

As hypothesized, people who received negative feedback were more likely to rationalize their performance by derogating the test, downplaying social sensitivity as a skill, and reporting that they were in a worse physical and psychological state. Also as hypothesized, they predicted that future negative feedback would not make them as unhappy as people in the other conditions said it would. The next step was to see if the changes in rationalization mediated the changes in predictions. First, we averaged the three rationalization scales (ratings of the test, importance of social aptitude, ratings of current state) to create an overall index of rationalization. Second, because the effect of negative feedback changed people's predictions for similar and dissimilar negative situations (see Figure 1), we averaged the predicted happiness ratings over both types of situations to create an index of predicted happiness. Third, because we predicted that positive feedback would have no effect on rationalization or predictions about future negative situations, we combined the positive and no feedback group into a control condition (dummy-coded 0) and compared them to the negative feedback group (dummy-coded 1). We then conducted a mediation analysis on these variables (see Kenny, Kashy, & Bolger, 1998, for details). The effect of negative feedback on the mediator, rationalization, was significant ($\beta = .38$, $SE = .09$, $p < 001$). The relationship between rationalization and predicted happiness also was significant, after controlling for feedback ($\beta = .22$, $SE = .10$, $p = .03$). The direct effect of negative feedback on predicted happiness was significant ($\beta = .32$, $SE = .09$, $p = .001$), but was reduced significantly when the effect of rationalization was controlled ($\beta = .23$, $SE = .10$, $p = .02$, $z = 2.00$, $p < .05$). These results meet all of the conditions necessary to demonstrate that changes in rationalization mediated the effects of feedback on predicted happiness.

In sum, negative experiencers generalized to new, negative situations, whereas positive experiencers did not generalize to new, positive situations. Interestingly, negative experiencers generalized even to situations that were not very similar to the test they had just taken, such as overhearing a stranger say they were not very sensitive. People appear to have rationalized the negative feedback rather broadly, inferring not only that the V-SAS was an invalid test but that the trait of social aptitude was not very important. As a result, they predicted that receiving a critical comment about their sensitivity would not make them as unhappy as people in the positive feedback and no feedback conditions predicted it would.

If this interpretation is correct, then it should be possible to limit the scope of people's generalizations by narrowing the kinds of rationalizations they make for their poor performance. We tested this hypothesis in Study 2 by attempting to get some participants in the negative feedback condition to narrow their rationalizations to the invalidity of the V-SAS test. We predicted that these participants would moderate their predictions about how unhappy they would be after receiving negative feedback on future V-SAS tests but not their predictions about how unhappy they would be after receiving negative feedback in the dissimilar situations (e.g., overhearing a negative comment from a stranger).

STUDY 2: NARROWING THE SCOPE OF PEOPLE'S RATIONALIZATIONS

Method

OVERVIEW

We replicated the no feedback and negative feedback conditions in Study 1 with the addition of a manipulation designed to narrow the scope of people's rationalizations for negative feedback. The study thus employed a 2 (feedback: negative or none) × 2 (rationalization: specific vs. undirected) × 2 (similarity of future situations: similar vs. dissimilar) design. We expected to replicate Study 1 in the undirected rationalization condition; negative experiencers should moderate their forecasts for how unhappy they would be in both similar and dissimilar future situations. Negative experiencers in the specific rationalization condition were expected to moderate their forecasts only for how unhappy they would be in the similar situations.

PARTICIPANTS

The participants were 127 undergraduates enrolled in psychology courses at the University of Virginia who received course credit for their participation. Of these, 111 met the criteria for the study of scoring 10 or less on the BDI (74 women and 37 men).

PROCEDURE

The negative and no feedback conditions of Study 1 were replicated exactly except for the following changes: After they took the V-SAS, half of the participants were randomly assigned to the specific rationalization condition. They received a questionnaire whose purpose was supposedly to obtain critical feedback on the V-SAS. Its true purpose was to suggest to people in the negative feedback condition that their poor performance was due to the problems with the test. The first set of questions suggested that the poor quality of the photographs might have made it difficult to judge people's emotions. The instructions noted that there were two versions of the test: one that used "small, black-and-white photocopies of photographs that are somewhat grainy" and one that used "high-quality, 5- × 7-inch color prints." The first question asked participants which set of photographs they had received (in fact, everyone had received the black-and-white photos). The second question asked how easy it was to detect people's emotions from the pictures they received. The third question noted that "some people have complained that it is especially difficult to detect blends of two or more emotions in the low-quality, black-and-white pictures" and asked people how easy it was for them to detect blends. People responded on 9-point scales (1 = *very difficult*, 9 = *very easy*). Three additional questions focused people's attention on other possible problems with the test; namely, it was too short, it used a multiple-choice format, and social aptitude is something that cannot be measured on a test. People responded to these questions on 9-point scales (1 = *disagree*, 9 = *agree*). People's ratings on these measures were reverse coded such that high scores reflected negative evaluations of the test. Participants were given 5 minutes to fill out the rationalization questionnaire. Those in the undirected rationalization condition sat quietly for 5 minutes.

Dependent measures. The dependent measures were identical to Study 1 except that we did not include the Associated Reasoning Scale. The prediction items were the same as in Study 1 except that people only rated the negative outcomes (e.g., getting a D on the tests or overhearing negative remarks). As a measure of how people construed the new situations, people also rated how valid each test or situation was (1 = *not very valid*, 9 = *extremely valid*) and how important the test or situation would be to them (1 = *not very important*, 9 = *extremely important*). People then were asked to recall what grade they had received on the V-SAS (all participants recalled their grade accurately), completed the same rationalization questionnaire as in Study 1, and were fully debriefed.

Results and Discussion

The baseline happiness ratings in the four cells of the Feedback × Rationalization design did not differ significantly, Fs(1, 106) < 1. As in Study 1, we subtracted people's baseline happiness ratings from their happiness ratings 5 and 15 minutes after taking the test.

EFFECTS OF FEEDBACK ON ACTUAL HAPPINESS

Once again, the negative feedback had little or no effect on people's happiness, as indicated by the absence of any significant effects in a 2 (feedback: none vs. negative) × 2 (rationalization: specific vs. undirected) × 2 (time: happiness ratings 5 vs. 15 minutes after the test) mixed-model ANOVA, Fs(1, 104) < 1.70, ps > .19. At 5 minutes, the means were –.07 and –.04 (SDs = .60, .76) in the no feedback and negative feedback conditions; at 15 minutes, the means were –.05 and .02 (SD s = .84, .84).

PREDICTIONS ABOUT FUTURE REACTIONS TO SIMILAR AND DISSIMILAR NEGATIVE EVENTS

As in Study 1, we averaged people's happiness predictions for the five tests that were similar to the V-SAS they had just taken (α = .95) and for the four situations that were dissimilar to the V-SAS (α = .91), and analyzed these predictions with a 2 (feedback: none vs. negative) × 2 (rationalization: specific vs. undirected) × 2 (similarity: similar vs. different future events) mixed-model ANOVA. As hypothesized, the three-way interaction was significant, $F(1, 106) = 4.09$, $p = .046$.

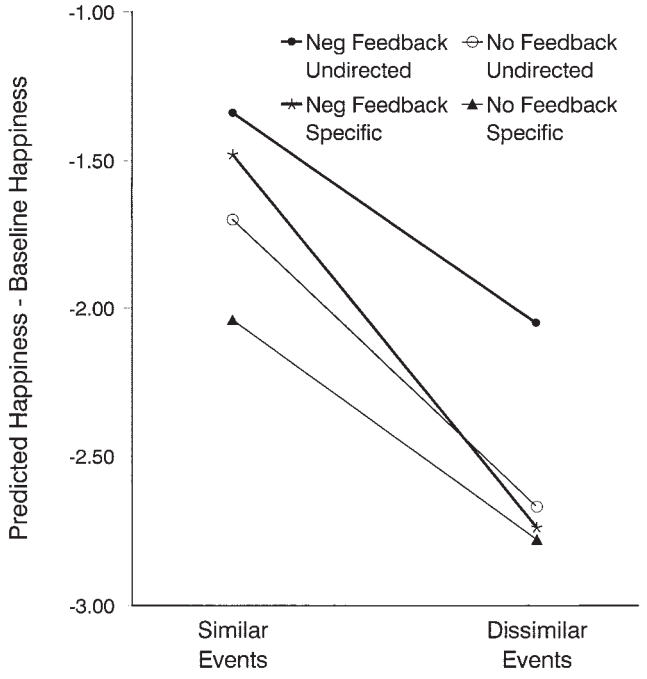

Figure 2 Study 2: Predicted happiness following similar and dissimilar future events.

As seen in Figure 2, the results in the undirected rationalization condition were similar to the results of Study 1. Negative experiencers again made more moderate predictions for both similar and dissimilar situations, compared to people in the no feedback condition. Although a 2 (actual feedback: negative vs. no feedback) × (future event: similar vs. dissimilar) ANOVA in the undirected rationalization condition revealed that the main effect of feedback was not significant, $F(1, 55) = 1.44$, $p = .24$, this main effect was significant when combined with the identical conditions in Study 1 ($z = 3.06$, $p = .002$). The reliability of this effect did not differ significantly across studies, $\chi^2(1) = 1.91$, $p = .17$.

As hypothesized, people in the negative feedback–specific rationalization condition appear to have made more limited generalizations. Compared to their counterparts in the no feedback condition, they predicted that similar negative experiences would not make them as unhappy but that dissimilar negative experiences would make them just as unhappy. The Feedback × Similarity of Event interaction was nearly significant in the specific rationalization condition, $F(1, 51) = 3.81$, $p = .057$.

RATIONALIZATION

In the specific rationalization condition, people evaluated the test right after taking it. If the manipulation was successful, then negative experiencers should have taken this opportunity to be more critical of the test than people who did not receive any feedback as a means of rationalizing their poor performance. This was the case; negative experiencers had significantly more negative responses on these questions than people who received no feedback (*M*s = 6.86 vs. 5.83, *SD*s = 1.22 and .99), $F(1, 50) = 11.31$, $p < .001$. In other words, the manipulation did not make all participants more critical of the test. Rather, people who received negative feedback were more likely to follow the lead of the questions and rate the test negatively.

At the end of the study, participants in all conditions completed the rationalization questionnaire used in Study 1. A factor analysis yielded a very similar solution (see Note 4). As in Study 1, negative experiencers rated the specific test more negatively (*M*s = 6.13 vs. 4.62, *SD*s = 1.35 and 1.28), $F(1, 105) = 35.84$, $p < .001$; they tended to say that social aptitude was a less important skill (*M*s = 2.89 vs. 2.44, *SD*s = 1.57 and 1.29), $F(1, 105) = 2.59$, $p = .11$, and that they were in a worse psychological and physical state while taking the test (*M*s = 3.55 vs. 2.97, *SD*s = 1.57 and 1.23), $F(1, 105) = 4.57$, $p = .04$.

There were no significant main effects or interactions involving the rationalization manipulation on these indices. As hypothesized, however, negative experiencers in the undirected and specific rationalization conditions both derogated the specific test, but only those in the undirected condition derogated social aptitude as a skill. On the latter measure, negative experiencers in the undirected condition derogated social aptitude as a skill somewhat more than people in the no feedback condition (*M*s = 6.91 vs. 7.57, *SD*s = 1.83 and 1.30), $F(1, 106) = 2.94$, $p = .09$ (this replicates the results of Study 1). In the specific rationalization condition, there was less of a tendency among negative feedback participants to derogate social aptitude as a skill (*M*s = 7.32 vs. 7.57, *SD*s = 1.22 and 1.29), $F(1, 106) < 1$. A planned contrast that weighted the negative feedback/undirected rationalization –3, and the other means 1, was nearly significant, $F(1, 105) = 3.21$, $p = .07$.

People also rated the validity and importance of each of the future situations about which they made predictions. Because the validity and importance ratings were highly correlated (.86), we averaged them. Consistent with the hypothesis that negative experiencers rationalized their poor performance, this group rated the similar situations (the V-SAS tests) as less valid and important than did people in the no feedback condition (*M*s = 3.18 vs. 3.90, *SD*s = 1.26 and 1.09). They also rated the dissimilar situations as less valid and important than did people in the no feedback condition (*M*s = 4.68 vs. 4.88, *SD*s = 1.33 and 1.15), but to a lesser degree, as reflected by a significant Feedback × Similarity of Events interaction, $F(1, 107) = 4.72$, $p = .03$.

The three-way interaction with the rationalization factor was not significant, $F(1, 107) < 1$. As predicted, however, there was some evidence that people in the negative feedback/undirected rationalization condition were more likely to rate the dissimilar situations as invalid and unimportant ($M = 4.40$, $SD = 1.33$) than did people in any of the other three conditions; means in the no feedback/undirected, no feedback/specific, and negative feedback/specific conditions were 4.87, 4.89, and 4.98, respectively (SDs = .95, 1.35, 1.29, respectively), $F(1, 107) = 3.57$, $p = .06$. This result is consistent with the idea that people in the negative feedback/undirected condition rationalized their poor performance rather broadly, leading them to rate even the dissimilar situations as less valid and important than people in the other conditions.

MEDIATION ANALYSES

As in Study 1, we tested whether rationalization mediated the effects of negative feedback on predicted happiness. First, we examined whether the mediation results of Study 1 replicated by performing the identical analysis in the undirected rationalization condition (which was equivalent to the same condition as run in Study 1). The results closely paralleled those of Study 1. The effect of negative feedback on rationalization was significant ($\beta = .54$, $SE = .11$, $p < 001$). The relationship between rationalization and predicted happiness also was significant, after controlling for feedback ($\beta = .37$, $SE = .15$, $p = .02$). The direct effect of negative feedback on predicted happiness ($\beta = .19$, $SE = .13$, $p = .16$) was reduced significantly when the effect of rationalization was controlled ($\beta = -.01$, $SE = .15$, *ns*, $z = 2.22$, $p < .05$). The only part of this analysis that did not replicate Study 1 is that the effect of negative feedback on predicted happiness was not significant. As mentioned earlier, however, this effect is significant when averaged across Studies 1 and 2.

Because we manipulated how people rationalized, there were more specific mediation hypotheses that we could test. First, we predicted that negative experiencers in the undirected and specific rationalization conditions would both moderate their predictions about how they would feel after doing poorly on similar tests because both had rationalized their poor performance by derogating these tests. To test this prediction, we dummy coded negative experiencers in both rationalization conditions 0 and no feedback participants 1 and performed mediation analyses of their predictions about similar events (the left-hand side of Figure 2). As predicted, the effect of negative feedback on rationalization was significant ($\beta = .43$, $SE = .09$, $p < .001$), as was the relationship between rationalization and predicted happiness, after controlling for feedback ($\beta = .29$, $SE = .10$, $p < .01$). Furthermore, the direct effect of negative feedback on predicted happiness ($\beta = .17$, $SE = .10$) was reduced significantly when the effect of rationalization was controlled ($\beta = .04$, $SE = .10$, *ns*, $z = 2.51$, $p < .05$).

When making predictions about dissimilar events, negative experiencers in the specific rationalization condition were expected to behave similarly to no feedback participants because they had denigrated the test rather than focusing on more general aspects of the situation. To test this hypothesis, we combined this group with no feedback participants (dummy coded 0) and compared them with negative experiencers in the undirected rationalization condition in an analyses of their predictions about dissimilar events (the right-hand side of Figure 2). As predicted, the effect of negative feedback on rationalization was significant ($\beta = .30$, $SE = .09$, $p < .001$), as was the relationship between rationalization and predicted happiness, after controlling for feedback ($\beta = .29$, $SE = .10$, $p < .005$). Furthermore, the direct effect of negative feedback on predicted happiness ($\beta = .21$, $SE = .10$) was reduced significantly when the effect of rationalization was controlled ($\beta = .12$, $SE = .10$, *ns*, $z = 2.29$, $p < .05$).

GENERAL DISCUSSION

Summary of Studies 1 and 2

When left to their own devices (i.e., when we did not attempt to direct people's rationalizations), negative experiencers predicted that future negative feedback would not make them as unhappy as did people who received no feedback ($z = 3.06$, $p = .002$) (averaged across studies). Negative experiencers generalized more to similar than dissimilar situations, as indicated by a reliable Feedback × Similarity of Situations interaction ($z = 2.01$, $p = .04$). Nonetheless, the feedback effect was significant for both similar and dissimilar situations (zs = 3.10 and 2.65, ps = .002 and .008).

The undirected negative experiencers rationalized their poor performance rather broadly. They devalued the specific test they had taken more than did people in the no feedback condition ($z = 5.86$, $p < .001$) and said that social aptitude was a less important skill ($z = 2.57$, $p = .01$). There was also a trend for negative experiencers to say that they were in a worse psychological and physical state while taking the test ($z = 1.83$, $p = .07$). This broad pattern of rationalization apparently led negative experiencers to believe that future negative feedback would not affect them as much in several different situations. Study 2 showed that people who were induced to make more specific rationalizations about the causes of their poor performance generalized more narrowly when making affective forecasts about the future. Furthermore, mediation analyses in both studies were consistent with the hypothesis that changes in rationaliza-

tion mediated the effects of negative feedback on predicted happiness.[5]

Thus, there is good evidence that negative experiencers moderated their forecasts about their reactions to future negative events because of the way in which they rationalized their poor performance on the test, consistent with our reconstrual hypothesis. As mentioned earlier, however, it is possible that there was also some genuine learning from experience; that is, the reconstrual hypothesis and learning-by-experience hypotheses could both be true. In addition to basing their forecasts on their reconstrual of the test, people might also, at least to some extent, have based their forecasts on their memory that the previous test did not influence them very much. If so, then there should be a correlation between people's actual happiness after the test and their predicted happiness to future negative events, more so than in the positive or no feedback conditions. This is because the actual reactions of people who received negative feedback are diagnostic for how they might feel after future negative feedback, whereas the actual reactions of people who received positive or no feedback are not.

The correlational evidence for the learning-from-experience hypothesis was mixed. In Study 1, the correlation between people's actual happiness 5 minutes after the test and their predicted happiness 5 minutes after future negative feedback was .62 in the negative feedback condition ($df = 35$, $p < .05$) versus .31 and .22 in the no feedback and positive feedback conditions, respectively (dfs = 34 and 31, *ns*). The former correlation was significantly higher than the average of the latter two ($z = 2.12$, $p < .05$). However, this evidence was weaker when the covariances were examined, which are less susceptible to condition differences in means and standard deviations. Furthermore, this finding did not replicate in Study 2, where the correlation was actually higher in the no feedback than the negative feedback condition (.45 vs. .24, dfs = 53 and 51), albeit nonsignificantly so ($z = 1.17$, *ns*). (There was no positive feedback condition in Study 2.[6]) Thus, the evidence for the learning-from-experience hypothesis is, at best, weak. We can conclude that learning from experience is not a necessary condition for negative experiencers to moderate their forecasts because there was evidence that they rationalized their poor performance and that this rationalization mediated their predictions of their future happiness.

In contrast, people did not generalize at all from the positive experience of doing well on the test. Even though receiving positive feedback had a minimal impact on people's happiness, people predicted that doing well on an identical test in the future would make them quite happy, as happy as people in the no feedback condition said it would. The reason people failed to generalize from positive experiences, we suggest, is that they did not make the effort to consult their recent experiences with the V-SAS test when making their forecasts; that is, they failed to satisfy the mental effort criterion by making "top-of-the head" forecasts about how happy they would be in the future situations. Consistent with this view, there were no significant between-condition differences in Study 1 in the correlations between actual happiness 5 minutes after the test and how they predicted they would feel 5 minutes after future positive feedback (rs = .38, .10, and .20 in the positive feedback, no feedback, and negative feedback conditions, respectively; $df = 31$, 34, and 35, respectively; positive feedback vs. average of other two conditions, $z = 1.14$, *ns*). Again, the covariances showed even less of a pattern; they were .20, .07, .15 in the positive feedback, no feedback, and negative feedback conditions, respectively.

Admittedly, our evidence for the hypothesis that people in the positive feedback condition did not consult their earlier experience and instead used a top-of-the-head strategy is indirect. Our strategy was to satisfy two of the three conditions necessary to learn from experience (accuracy of recall and applicability) and to infer that if people still failed to generalize, they must not have met the third criterion (mental effort). Accuracy of recall was satisfied by asking people to make predictions soon after experiencing the event; presumably, people could remember how they had felt 15 minutes earlier. Applicability was satisfied by asking people to forecast their feelings for very similar events in the future, one of which was identical to the test they had just taken. The fact that people still failed to generalize from their recent experiences suggests that they did not engage in the mental effort necessary to compare their recent experiences with the future ones.

We do not mean to suggest that people will never engage in such mental comparison or that they will never learn from positive experiences. In the present studies, people may have used a top-of-the head strategy because the future events they were considering were hypothetical ones. When making more consequential forecasts about real events, people might expend the mental effort necessary to retrieve their previous experiences from memory and apply them to their forecasts about similar events in the future. Even if they did, however, they might not always succeed in recalling their past reactions accurately. In everyday life, people often think about future events (e.g., next month's vacation) long after their experience with similar events (last summer's vacation). Meyers et al. (2000) and Mitchell et al. (1997) found that after time passes, people exaggerate how happy they were during and after positive events. Thus, whereas the three conditions necessary to learn from positive experiences may sometimes be met, we suspect

they typically are not in everyday life. Such a conclusion explains why the durability bias for positive events has been found in many settings (Gilbert et al., 1998; Wilson et al., 2000), even in ones in which people had ample past experience.

Limited Learning From Experience

The present studies seem to imply that people will be more accurate at predicting their emotional reactions to negative than positive events, at least for negative events they have experienced in the past. We believe there are a number of reasons, however, to doubt this view. First, we have found ample evidence for a durability bias for negative outcomes in previous studies, even for negative events that people had experienced before and could have learned from (e.g., a loss by one's favorite football team; Gilbert et al., 1998; Wilson et al., 2000). Second, even though negative experiencers in Studies 1 and 2 moderated their forecasts to some extent, they still predicted that the future negative events would make them quite unhappy.

The studies were not designed to test the accuracy of people's forecasts. We did not bring people back to the lab and give them another bad grade or arrange for them to be insulted by a neighbor. In our research on affective forecasting, however, we have been struck by how quickly people recover from similar negative events. In the present studies, we found no evidence that receiving negative feedback on the V-SAS test made people unhappy. People who received negative feedback were no less happy 5 minutes later than people who received no feedback ($z = -.08$, *ns*). In addition, there were not any differences on an indirect measure of mood—the Associated Reasoning Test (Study 1). Thus, the fact that negative experiencers in Studies 1 and 2 predicted that a poor performance on a test 2 weeks later would lower their happiness by between 1 and 2 points on the happiness scale (see Figures 1 and 2) makes us suspicious. It is possible, of course, that doing poorly on a V-SAS test a second time would make them substantially unhappier than doing poorly on it the first time. Given how resilient we have found people to be after several different kinds of negative experiences, however, we suspect that people in the negative feedback conditions were overestimating how unhappy they would be, albeit to a lesser extent than were people in the no feedback or positive feedback conditions.

In short, there seems to be a paradox: We found that people who experienced a negative event predicted that future negative occurrences would not influence their happiness as much as people who did not experience the negative event. However, in previous studies and in the present ones, we have found ample evidence for a strong durability bias for predictions for negative events. Why do people appear to have learned from negative experiences but continue to commit the durability bias for negative events?

One possibility is that the magnitude of the durability bias for negative events is so large that it remains even after people have corrected it to some degree. People may have learned from their past negative experiences but not enough to completely avoid the durability bias. This interpretation explains why the bias continues to be as strong or stronger for negative events than for positive events. The two mechanisms known to produce the durability bias—focalism (viewing the event in a vacuum) and immune neglect (underestimating how much they will rationalize the event)—both operate on forecasts about negative events. Only one of these mechanisms (focalism) operates on forecasts about positive events, because people are not motivated to rationalize away positive experiences. Uncorrected forecasts, then, would be expected to be more extreme for negative events. Thus, even if people do learn from past negative experiences, the magnitude of the durability bias could still be quite large.

We close by noting that it may not always be advisable to correct the durability bias. Overestimating the duration of emotional reactions can serve as a motivator, energizing people to work for positive outcomes and avoid negative ones. We are not convinced, however, that the durability bias is an altogether good thing, given that many important decisions are based on forecasts about the duration of affective reactions, such as whom to marry, whether to become a lawyer or psychologist, and whether to plant daffodils in the garden. It is to people's advantage to know exactly which events really will cause lasting pleasure and which will not. If people knew that they will recover quickly if their favorite team loses the big game next week, they might spend less time obsessing about the game and more time planting daffodils.

NOTES

1. A durability bias, we should note, could result from at least two kinds of errors in prediction. People could overestimate the peak intensity of their initial emotional reaction, underestimate the rate at which they will recover from the reaction, or both. People could be wrong about how badly they will feel 30 minutes after getting a bad grade on a test, for example, because they overestimated how badly they felt right after getting the grade, underestimated how quickly they would recover from the negative feedback, or both. Although it is important to consider the conditions under which each of these errors occur, the purpose of the present studies was not to address this question but to investigate the conditions under which people learn about the durability bias more generally; that is, regardless of whether people overestimate initial intensity or underestimate recovery, after the event, they end up with a less intense emotional reaction than they predicted, and when and how they come to recognize this fact is the question of the present studies.

2. In both the Mitchell, Thompson, Peterson, and Cronk (1997) and Meyers, Wilson, and Gilbert (2000) studies, the retrospective durability bias was not as strong as the prospective one. Meyers et al. sug-

gested that focalism is probably weaker when recalling the past than when predicting the future. It is easier to "fill in the vacuum" of the past by recalling that other events influenced one's thoughts and feelings than to imagine the occurrence of future events.

3. Very similar results were found in both Studies 1 and 2 when baseline happiness was used as a covariate rather than performing an ANOVA on difference scores.

4. The same factor analysis was performed on the rationalization measures in Study 2 and a similar three-factor structure was found. Because exploratory factor analyses of small samples are unstable, we decided to keep only items that loaded .4 or more on their respective factor in both data sets. All items except the two mentioned in the text met this requirement.

5. These tests of significance averaged over the two studies using the method of adding *t*s (Rosenthal, 1978). In no case did the reliability of the effects differ significantly across studies, with one possible exception. There was a marginally significant tendency for the effects of negative feedback on predictions about similar situations to differ across studies, $\chi^2(1) = 3.42$, $p = .06$.

6. All of these correlations are between people's reported happiness 5 minutes after the test and their predicted happiness. The correlations between people's reported happiness 15 minutes after taking the test and their predicted happiness were very similar.

REFERENCES

Buehler, R., & McFarland, C. (2001). Extremity bias in affective forecasting: The role of temporal focus. *Personality and Social Psychology Bulletin, 27*, 1480-1493.

Christianson, S. A., & Safer, M. A. (1996). Emotional events and emotions in autobiographical memory. In D. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 218-243). New York: Cambridge University Press.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face.* Englewood Cliffs, NJ: Prentice Hall.

Erskine, A., Morley, S., & Pierce, S. (1990). Memory for pain: A review. *Pain, 41*, 255-265.

Festinger, L. (1957). *A theory of cognitive dissonance.* Palo Alto, CA: Stanford University Press.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*, 107-119.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75*, 617-638.

Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology, 46*, 44-56.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles.* New York: Guilford.

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

Koehler, J. J. (1996). The base rate fallacy reconsidered: Normative, descriptive, and methodological challenges. *Behavioral and Brain Sciences, 19*, 1-53.

Levine, L. J. (1997). Reconstructing memory for emotions. *Journal of Experimental Psychology: General, 126*, 165-177.

Mayer, J. D., & Hanson, E. (1995). Mood-congruent judgment over time. *Personality and Social Psychology Bulletin, 21*, 237-244.

Meyers, J. M., Wilson, T. D., & Gilbert, D. T. (2000). [The accuracy of predicted and recollected happiness after emotional events]. Unpublished raw data.

Mitchell, T. R., Thompson, L., Peterson, E., & Cronk, R. (1997). Temporal adjustments in the evaluation of events: The "rosy view." *Journal of Experimental Social Psychology, 33*, 421-448.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* New York: Springer-Verlag.

Rosenthal, R. (1978). Combining the results of independent studies. *Psychological Bulletin, 85*, 185-193.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341-357.

Ross, M., & Newby-Clark, I. R. (1998). Construing the past and future. *Social Cognition, 16*, 133-150.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*, 67-85.

Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology, 59*, 291-297.

Wilson, T. D., Wheatley, T. P., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 78*, 821-836.

Wong, P.T.P., & Weiner, B. (1981). When people ask "why" questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology, 40*, 650-663.

Received September 27, 2000

Revision accepted December 31, 2000