

Greener grass or sour grapes? How people value future goals after initial failure

Hallgeir Sjøstad
Center for Applied Research at NHH and Norwegian School of Economics

Roy F. Baumeister
University of Queensland

Michael Ent
Towson University

Abstract

Across six experiments (N=1,304), people dealt with failure by dismissing the value of future goals. Participants were randomly assigned to receive good or poor feedback on a practice trial of a cognitive test (Studies 1-3, 5-6) or their academic performance (Study 4). Those who received poor (vs. good) feedback predicted that they would feel less happy about a future top performance. However, when all participants received a top score on the actual test they became equally happy, regardless of initial feedback. That is, initial failure made people underestimate how good it would feel to succeed in the future. Inspired by Aesop's fable of the fox and the grapes, we term this phenomenon the "sour-grape effect": A systematic tendency to downplay the value of unattainable goals and rewards. Mediation analyses suggest that the low happiness predictions were a self-protective maneuver, indicated by apparent denial of the personal and future relevance of their performance. Moderation analysis showed that people high in achievement motivation constituted the main exception, as they predicted (correctly) that a big improvement would bring them joy. In a final and high-powered experiment, the effect generalized from predicted happiness to predicted pride and gratitude. Crucially, the sour-grape effect was found repeatedly across two different countries (USA and Norway) and multiple settings (lab, field, online), including two pre-registered replications. In line with the principle of "adaptive preferences" from philosophy and cognitive dissonance theory from psychology, the results suggest that what people want is restricted by what they can get.

Keywords: goals; happiness; cognitive dissonance; sour grapes; affective forecasting

Abstract word count: 248

Paper word count: 12,838

Introduction

Folk wisdom has been a rich but also frustrating source of insights for the social sciences, particularly because of its endorsement of contradictory predictions. For instance, research on interpersonal attraction has often invoked competing hypotheses based on whether “opposites attract” or “birds of a feather flock together”. The present investigation was inspired by another pair of competing hypotheses from folk wisdom, concerning how people feel about future rewards that seem out of reach.

The idea that people would respond to such rewards by devaluing them was expressed in Aesop’s classic fable about the fox and the grapes. After leaping unsuccessfully to reach a bunch of sweet grapes hanging too high above his head, the fox changed his mind and decided that they were probably sour anyway (e.g., Elster, 1983; Temple, Temple, & Temple, 1998). In contrast, the English proverb “the grass is always greener on the other side of the fence” proposes that people may consider out-of-reach rewards especially desirable. From that proverb’s perspective, the fox should have opined those remote grapes would be extra sweet, not sour. We note that this pair of opposing narratives may seem equally plausible, but they cannot both be correct at once under well-specified conditions. Informed by psychological theory and prior research, the current investigation subjected these competing tidbits of folk wisdom to a series of experimental tests.

In the present studies, we created expectancies by having participants perform an initial cognitive task for which they received either good or poor feedback. Participants were told that a second performance lay ahead, sometimes with a reward possible if they could achieve superior performance. Positive events that are unexpected or surprising have been shown to yield more joy and happiness than similar events that were expected (e.g., Shepperd & McNulty, 2002), which lends support to the greener-grass hypothesis. Indeed, economic principles and scarcity theories would seemingly all assert that the subjective value of a good

should increase as the resource becomes less available (i.e., price increases as supply drops relative to demand). For instance, based on persuasion research (e.g., Cialdini, 1993), it has become a common marketing strategy to reduce the perceived availability of products to create a sense of scarcity and thereby increase demand or willingness to pay. Applying this perspective to the present studies, people should expect and actually feel a greater sense of happiness about reaching the high goal if their first-round performance had been poor rather than good.

On the other hand, if people seek to evaluate future rewards so as to safeguard a positive sense of self, they might devalue the goal when it seems unlikely (i.e., after a poor first-round performance) — just like Aesop's fox. Indeed, cognitive dissonance theory (Festinger, 1957; Aronson, 1992) states that people make motivated adjustments to their attitudes in order to maintain a positive self-concept. Ego threats and other uncomfortable inconsistencies therefore cause people to change their attitudes strategically, which may also happen for anticipated outcomes (e.g., Arrowood & Ross, 1966; Shaffer & Hendrick, 1971; Ent & Gerend, 2015). This theory could propose that people would reduce cognitive dissonance associated with a futile effort by withdrawing emotional investment and reducing personal identification with the task at hand. In philosophy, this idea has been summarized in the principle of "adaptive preferences": The tendency for human desire to adapt to what is possible (Elster, 1983).

Several lines of work converge to make the sour-grape hypothesis plausible. Self-handicapping research (Berglas & Jones, 1978) showed that people respond to situations that contain subjectively hard-to-reach goals by withdrawing effort and seeking to protect themselves from the implications of failure, even at the cost of increasing the likelihood of failure. Shepperd, Findlay-Klein, Kwavnick, Walker, and Perez (2000) found that when feedback is imminent, people revise their expectancies downward, apparently in a strategic

move to protect themselves against possible disappointment. Focusing on the mental adjustment to uncontrollable events, one study found that participants who felt very certain that they had lost a dating game evaluated the date more negatively than participants who believed they still had a fair chance (Wilson, Wheatley, Kurtz, Dunn, & Gilbert, 2004). Other research has found that both favorable and unfavorable outcomes are rated as more desirable when their perceived likelihood increases, in the cases of both a presidential election and potential changes in tuition rates at a university (Kay, Jimenez, & Jost, 2002). Although none of these findings directly predicts how people would respond to the present situation, they have the common theme that people seek to avoid and minimize disappointment, and so they all might be congenial to the sour-grape hypothesis that people would devalue future rewards if they believe that they are unlikely to be gained.

Whether anticipatory devaluation would carry over into actually reducing the subjective enjoyment of success when it does occur is unclear. If one has devalued the success in advance, it may become less satisfying, or possibly not. Early work on the irreversibility of dissonance reduction processes suggests that devaluing the success would indeed make it less satisfying (Lepper, Zanna, & Abelson, 1970). In more recent research, one study found no effect of prior expectations on current emotion (Golub, Gilbert, & Wilson, 2009), whereas a second study found that optimistic expectations could lead to greater disappointment after a poor test result (Sweeny & Shepperd, 2010). Thus, the existing evidence is mixed. Put another way, in Aesop's tale, the fox slunk away leaving the grapes untouched, and so we do not know how much he would have actually enjoyed them — especially if he had tasted them after having pronounced them likely to be sour.

Happiness Forecasting as Reward Valuation

As a measure of reward valuation, the primary dependent variable in the present experiments is *happiness forecasting* — that is, how happy people predict they will feel if they were to succeed on a future task, after learning the final outcome. In our final experiment (Study 6), we used a broader measure of emotional well-being as an extended replication, consisting of predicted happiness, pride and gratitude (DeSteno, 2018).

From classic utility theory (Bentham, 1789) to modern decision research (e.g., Bruni & Sugden, 2007; Read, 2007), the value of any action or outcome can be conceptualized as the total amount of pleasure and pain it brings (for review, see Morewedge, 2014). Hedonic value serves as a basic motivation for humans and nonhuman animals alike; wanting rewards and avoiding punishment are universal. However, while most or all nonhuman species seem to be “stuck in time,” to use Roberts’ (2002) term, humans are capable of mental time travel into both the past and the future (for review, see Roberts, 2002; Suddendorf & Corballis, 1997, 2007), thus enabling the person to go beyond immediate incentives. Decisions can also be guided by the *anticipated* value of future rewards. Future-oriented thinking at this level is rare in nature, but there is good evidence showing that it plays a large role in people’s everyday life and particularly in decision-making (Baumeister, Maranges, & Sjøstad, 2018). Therefore, we chose happiness forecasting as our primary outcome measure of reward valuation.

Since the 1970s, a substantial body of research has been devoted to happiness forecasting or “anticipated utility” (e.g., Brickman, Coates, & Janoff-Bulman, 1978; Kahneman & Snell, 1990, 1992; Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998; Mellers & McGraw, 2001; Loewenstein, O’Donoghue, & Rabin, 2003). A central finding is the so-called impact bias: People tend to overestimate their future emotions (for review, see Wilson & Gilbert, 2005). More precisely, people tend to predict the correct type of emotion

but overestimate its intensity and duration. For example, people expect lasting happiness following job promotions and lasting misery following romantic breakups, though the actual level of emotion tends to return to baseline sooner than predicted. These findings led Kahneman (2011, p. 402) to conclude that “nothing in life is as important as you think it is, while you are thinking about it.” However, recent theoretical work suggests that even if happiness forecasts are systematically inaccurate, the impact bias might have an adaptive function to motivate future-directed behavior (Miloyan & Suddendorf, 2015). Indeed, a meta-analysis found that anticipated emotions have a stronger effect on judgments and behavior than experienced emotions in the present (DeWall, Baumeister, Chester, & Bushman, 2016). In a set of experimental studies, Morewedge and Buechel (2013) found that overestimated emotions were more motivating than less extreme forecasts, leading to higher effort in goal pursuit.

Importantly, none of the aforementioned studies directly addressed how people would value future goals after a poor first performance. Hence, the present study builds on and extends the happiness forecasting literature by asking a different question: After *initial failure*, will the anticipated value of a future success tend to increase (greener grass) or decrease (sour grapes)? Moreover, less is known about the boundary conditions and underlying mechanisms of goal devaluation, and what self-protective strategies people may apply when predicting their future well-being. We therefore included established trait measures of fear of failure and achievement motivation as potential moderator variables, and the personal and future relevance of the given performance as potential mediator variables.

The Present Investigation

In the present research, we tested the competing speculations of greener grass and sour grapes by reformulating them as testable, theory-driven hypotheses. We experimentally manipulated

performance feedback in a trial run, thereby creating favorable or unfavorable expectancies of the upcoming trial. Participants then predicted how happy they thought they would be if they achieved success and got a top score, which enabled us to test whether participants with poor first-round feedback would devalue the reward (sour grapes) or value it all the more (greener grass). In two of the studies, participants actually performed this second task and received excellent feedback later in the experiment, so we could measure whether their predictions were accurate. In a conceptual extension we dispensed with manipulated feedback and created similar expectancies simply by reminding some participants of their current grade-point average before measuring their predictions of how happy they would be to graduate with a lofty A average.

The greener-grass hypothesis predicted that people should value hard-to-obtain goals more than more attainable goals. The sour-grape hypothesis predicted the opposite — namely, that people should devalue the goal when they believe that it is probably out of reach.

We report converging evidence from six experiments combining laboratory, field and online studies, conducted in two different countries with 1304 participants in total. To establish robustness and reproducibility, Study 5 was an incentivized and pre-registered replication of the key findings in Studies 1–3. Study 6 was a high-powered extended replication, which included predicted pride and gratitude in a broader outcome measure. We end by summarizing the results across all experiments using internal meta-analysis. The sample size in our final replication was informed by the observed effect in Studies 1-5, the " $N = x \times 2.5$ heuristic" for high-powered replication studies (Simonsohn, 2015), and a sensitivity power analysis. All measures, manipulations and exclusions are reported in each study.

We chose to include only two conditions in each experiment, creating a high versus low level of the independent variable (i.e., poor or good test feedback). Given that statistical power is a function of both sample size per cell and the expected effect size (and hence the

strength of the experimental manipulation), we designed our studies this way to maximize the probability of detecting a true and robust effect if it should exist. Seen as a whole, we found strong support for the sour-grape hypothesis: A systematic tendency to downplay the hedonic value of unattainable goals and rewards.

Study 1: Lab-in-the-Field Experiment

We conducted Study 1 as an experiment in a field setting, located in a city center in Norway. Following a cognitive test of “intuitive intelligence”, participants were told that their initial performance was either good or bad, using a false-feedback paradigm. The underlying assumption was that this variation would create the impression that participants were either more or less able to get a top score on the second test. At this point, they predicted how happy they would feel if that should happen. After giving participants manipulated feedback following the second test as well, all participants received a top score before we measured their actual happiness level. This enabled us to test the causal effect of poor versus good test performance on happiness predictions, and also to compare these predictions with their actual experiences after receiving a top score on the second test.

The two tests used items from the Cognitive Estimation Test (Shallice & Evans, 1978; Fein, Gleeson, Bullard, Mapou, & Kaplan, 1998), which requires people to estimate and extrapolate (e.g., “How many seeds are there in a watermelon?”). It is nearly impossible to know precise correct answers, so the experimenter said that getting within 10% would be counted as correct. This made the false feedback manipulation relatively immune to suspicion.

If the greener-grass hypothesis is true, participants receiving poor feedback should predict more happiness than participants receiving good feedback. If the sour-grape hypothesis is true, however, the opposite pattern should occur: Participants receiving poor (vs.

good) feedback on the first test should predict less happiness following a top score on the second test.

Method

Participants. We recruited 82 Norwegian participants (44 women; age: $M = 29.5$ years) in the field in exchange for a candy bar or a bottle of fruit juice (their choice). The experiment was conducted in city locations with calm surroundings, such as parks and other walkable areas without car traffic.

Materials and Procedure. Participants were randomly assigned to either a good or a poor feedback condition in a between-subjects design with two conditions. Initially, everyone was asked to participate in the validation of a novel test of “intuitive intelligence.” The experimenter introduced the test by saying that intuitive intelligence is “the ability to recognize patterns and solve problems quickly”, and explained further that “it is an important part of intelligence that some people have more of than others” and “has proven to be strongly related to positive outcomes in life, such as extraordinary achievements in work and a well-functioning social life.” Finally, the experimenter said that when the validation study was completed, the test would be used to select top managers and military pilots. Participants were informed that they would take two brief versions of this test.

We constructed two six-item tests, taking items from the Cognitive Estimation Test (e.g., “How much does 20 average sized apples weigh?”). The experimenter said that people with high intuitive intelligence have very reliable gut feelings, which enable them to provide quick and “accurate enough” answers based on logic and prior knowledge. They were also told that all answers within the range of $\pm 10\%$ would be evaluated as correct. The experimenter used a stopwatch during the testing and told participants that they had to

respond to each question within 20 seconds. By random assignment, half the participants were told that they had gotten 5 out of 6 correct, which was indicative of very good intuition. The rest were told they had only gotten 2 out of 6 correct, which was indicative of poor intuition.

The first dependent variable was measured here, in which the key question asked how happy the participant would feel if he or she were to receive a top score on the second test: “If you manage to get 6/6 correct on the second test of intuitive intelligence, how happy would that make you feel?” (0 = “not at all happy,” 10 = “very happy”). As a manipulation check, participants were also asked how difficult they thought it would be to get 6/6 correct (0 = “not at all difficult,” 10 = “very difficult”). After finishing the second test with a different set of 6 questions from the Cognitive Estimation Test, all participants were given manipulated feedback again, which was that they had gotten all 6 items correct (regardless of condition), indicating excellent intuition. To measure the second dependent variable, they were asked to report their current level of happiness: “In general, how happy do you feel right now?” (0 = “not at all happy,” 10 = “very happy”).

At the end of the study, all participants responded to an additional manipulation check that assessed how difficult they thought the first and second tests were in retrospect (0 = “not at all difficult,” 10 = “very difficult”). We did this to ensure that perceived difficulty was a function of the test feedback they received. Finally, all participants were fully debriefed, probed for hypothesis guessing, and thanked for their participation.

Results

Manipulation Checks. Participants in the poor feedback condition predicted that it would be more difficult to get a top score than participants in the good feedback condition ($M = 8.20$, $SD = 1.65$ vs. $M = 6.68$, $SD = 1.99$; $t(80) = 3.75$, $p < .001$, $d = .83$). Thus, the manipulation was effective at creating differences in perceived difficulty of success.

Furthermore, participants who got low scores on the first test rated it as more difficult than those who got high scores ($M = 7.05$, $SD = 1.53$ vs. $M = 5.66$, $SD = 2.14$; $t(80) = 3.38$, $p = .001$, $d = .75$). This may seem counterintuitive insofar as the two groups took exactly the same test and presumably exerted the same effort. However, a test should be considered more difficult when identical performances yield lower rather than higher grades. Consistent with this, the two groups' ratings of the second test, on which everyone received a perfect score, showed no difference in perceived difficulty ($M = 5.24$, $SD = 1.81$ vs. $M = 5.32$, $SD = 2.13$; $t(80) = .16$, $p = .87$, $d = .03$).

Predicted Happiness. In line with the sour-grape hypothesis, an independent t -test revealed that participants in the poor feedback condition predicted that they would feel less happy after receiving a top score on the second test than participants in the good feedback condition ($M = 5.46$, $SD = 1.55$ vs. $M = 6.41$, $SD = 1.73$). The difference was statistically significant, and the estimated effect size was of medium strength ($t(80) = 2.62$, $p = .011$, $d = .58$).

Actual Happiness. The level of actual (experienced) happiness following a top score on the second test was nearly identical between conditions ($M = 7.17$, $SD = 1.72$ vs. $M = 7.10$, $SD = 2.10$; $t(80) = .17$, $p = .860$, $d = .04$). Next, each participant's predicted level of happiness was subtracted from their actual self-reported happiness after the test, and we analyzed the difference scores. Participants in the poor feedback condition underestimated their future happiness to a greater extent than participants in the good feedback condition ($M = 1.71$, $SD = 2.12$ vs. $M = 0.68$, $SD = 1.72$). The difference was significant, and the effect size was of medium strength ($t(80) = 2.40$, $p = .019$, $d = .53$). See figure 1 for visual illustration of the results.



Fig. 1. Study 1 (N=82): In line with the sour-grape hypothesis, participants in the poor feedback condition predicted less happiness following a future top score than participants in the good feedback condition ($p = .011$, $d = .58$). However, there was no difference in actual happiness after receiving a top score on the second test ($p = .86$, $d = .04$). Error bars indicate standard error. Both outcome variables were measured using a 0–10 response scale.

Discussion

Contrary to the proverb of "greener grass", Study 1 found a pattern that were consistent with the sour-grape hypothesis: Participants who were randomly assigned to receive poor feedback on the first test predicted lower happiness in reaction to a future perfect score than participants who received good feedback. However, when all participants received a top score on the second test, the actual happiness level was indistinguishable between the two conditions. In our interpretation, this suggests that believing that a perfect score was unattainable made people devalue the goal to protect the self from failure and disappointment. Fortunately for the participants, the forecast of low happiness did not appear to detract from their actual joy when they achieved the lofty goal after an initial failure.

The remaining studies (2-6) sought to replicate the sour-grape effect in larger samples, while testing for generalizability and boundary conditions across different methods, settings and languages. Furthermore, we investigated the role of mediating process mechanisms and moderating trait characteristics, to better understand the underlying psychology of the effect.

Study 2: Online Experiment

Study 2 was conducted online, using a Stroop test as the basis for the false-feedback procedure. The test was presented as a reaction-time measure of the (fictitious) cognitive trait “mental agility.” As in Study 1, this cognitive ability was presented as a valuable trait that was strongly related to intelligence, success, and accomplishments in life. These changes help ensure that the effect is generalizable across contexts and different conceptualizations of cognitive ability (familiar and unfamiliar).

More importantly, Study 2 began our exploration of relevant mental processes, by including a measure of identity relevance as a possible mediator. If people want to defend themselves against implications of possible failure, they can do so by saying that this particular ability is not relevant to who they are as a person. This would help explain their prediction that even a perfect score would not bring them much joy. In other words, success is presumably less gratifying when it is in relation to something that is irrelevant to one’s sense of self, as opposed to succeeding on something more central to the self.

Based on the sour-grape hypothesis, we offer three predictions in this study. First, we predicted that Study 2 would replicate Study 1’s finding of reduced happiness forecasting following poor feedback. Second, we predicted that participants who received poor feedback would rate the test as less relevant to their self-concept, as compared to participants who received a good score. Third, we predicted that lowered identity relevance following poor feedback would mediate the negative effect on happiness forecasting.

Method

Participants. We recruited 200 American participants (138 women; age: $M = 39.5$ years) from an online sample at Amazon Mechanical Turk. They participated in a brief study of “personality and reaction-time” in exchange for \$0.40.

Materials and Procedure. Participants were randomly assigned to either a good or a poor feedback condition in a between-subjects design with two conditions. Unlike Study 1, the prospect of a future test was presented as occurring a year from now rather than in a few minutes. Since personal improvement and ambitious goal achievement usually take time to unfold, we made this change to make the prediction scenario appear more realistic. In addition, a measure of identity relevance was included as a mediator variable, indicative of cognitive dissonance as process mechanism. The logic of the feedback procedure was the same as in Study 1.

Initially, everyone was asked to participate in the validation of a novel test of “mental agility,” which was described as a highly beneficial psychological trait. Participants completed a 50-trial Stroop test and received manipulated feedback that their level of “mental agility” was either very good (top 20% in the population) or very poor (bottom 20% in the population).

After receiving test feedback, participants rated the identity relevance of their level of cognitive ability on a sliding scale ranging from 0 (“not at all”) to 100 (“very much”): “To what extent does your level of mental agility reflect who you are as a person?” Next, they predicted how happy they would feel if they were to achieve a top performance on a similar test in the future: “Imagine that in one year from now, you manage to get a top score on a similar test of mental agility (top 10% in the population). How happy would that make you feel?” (0 = “not at all happy,” 100 = “very happy”). Last, all participants were fully debriefed and thanked for their participation.

Results

Identity Relevance. Participants in the poor feedback condition indicated much lower identity relevance of their cognitive ability level than participants in the good feedback

condition ($M = 36.17$, $SD = 31.01$ vs. $M = 71.49$, $SD = 23.09$). The difference was statistically significant, and the estimated effect size was very large ($t(198) = 9.14$, $p < .001$, $d = 1.29$).

Predicted Happiness. In line with the sour-grape hypothesis, an independent t -test revealed that participants in the poor feedback condition predicted that achieving top performance on a similar cognitive test a year from now would make them less happy than what participants in the good feedback condition predicted ($M = 76.78$, $SD = 23.64$ vs. $M = 85.51$, $SD = 17.96$). The difference was statistically significant, and the effect size was moderate to medium ($t(198) = 2.94$, $p = .004$, $d = 0.42$).

Mediation Analysis. The results from a simple mediation analysis in PROCESS (Model 4) was consistent with a model in which identity relevance mediates the effect of negative test feedback on happiness forecasting.

Specifically, poor (vs. good) feedback led to lower identity relevance of cognitive ability level ($a = -35.31$, $p < .001$), and lower identity relevance was in turn associated with lower happiness forecasts following a future top performance ($b = .27$, $p < .001$). This indirect effect ($ab = -9.66$) was statistically significant, as a bias-corrected bootstrap confidence interval based on 10,000 bootstrap samples did not include the value 0 ($CI = -15.04, -4.97$). The total effect of test feedback on happiness forecasting was significant ($c = -8.73$, $p = .004$), while there was no evidence of a direct effect independent of the identity relevance mechanism ($c' = .93$, $p = .779$).

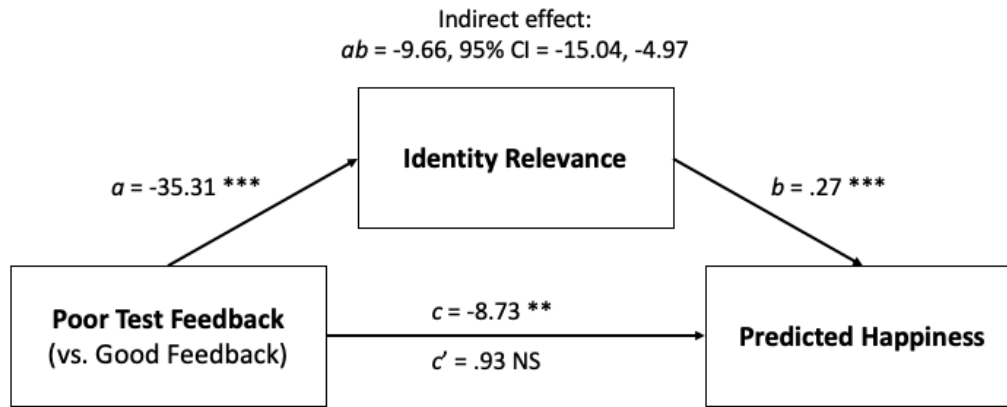


Fig. 2. Statistical mediation diagram for Study 2 (N=200), showing path values as unstandardized regression coefficients. Receiving poor feedback led to lower identification with one's cognitive performance than receiving good feedback. This reduced identification mediated the negative effect of poor test feedback on predicted happiness following a future top performance.

Discussion

Despite several minor changes in procedure, Study 2 replicated the finding from Study 1:

Personal predictions of future happiness conformed more to the sour-grape hypothesis than the greener-grass hypothesis. Participants who were told that they scored relatively poorly on the first test predicted that a future success on the same test (even one year from now) would bring them relatively little happiness, as compared to participants who were told that their first performance had been quite good.

Study 2 also provided evidence of defensive motivation as a process mechanism. Participants who scored poorly on the first test rated it as less relevant to their self-concept than those who scored well. Moreover, these changes in identity relevance mediated the predictions of future happiness. In effect, people reacted to initial failure by saying “This test is not relevant to me, and therefore I would not care even if I were to perform quite well next time.”

Study 3: Laboratory Experiment

Study 3 attempted to replicate the sour-grape effect in a controlled laboratory experiment, while extending the previous findings with two novel aspects. In addition to the measure of

identity relevance from Study 2, we included an additional mediator variable also indicative of cognitive dissonance, namely, to what extent the cognitive ability in question was deemed important in determining one's future success in life. Furthermore, we changed the outcome measure from happiness forecasting (how happy participants would feel after a future top performance) to perceived importance of the cognitive ability for one's future happiness.

The sour-grape hypothesis would predict that participants who received poor feedback would rate the ability as being largely irrelevant to their future happiness, compared with those who received good feedback. Self-concept defense would be indicated if participants who received poor feedback rated the test as irrelevant to their self-concept and as irrelevant to their future success in life, and if these responses mediated thinking that the cognitive ability being tested was irrelevant to their future happiness.

Method

Participants. We recruited 103 participants (73 women; age: $M = 19.75$ years) from a psychology course at an American university for a study of "personality and reaction-time" in exchange for course credit.

Materials and procedure. Each participant was randomly assigned to one of two conditions in a between-subjects design (good feedback vs. poor feedback). As in Study 1, the experimenter introduced the test by explaining the inherent value in having a high level of "intuitive intelligence", emphasizing its advantages both in work settings and in social life. The participants were then provided with a test consisting of 10 questions derived from the Cognitive Estimation Test, similar to Study 1. They were given 20 seconds per question before the screen automatically advanced to the next question. As in Study 1, participants were told that all answers within the range of $\pm 10\%$ would be evaluated as correct. After

completing the test, participants received manipulated feedback that they either performed well (top 20% in the population) or performed poorly (bottom 20% of the population).

At this point, all participants answered the questions that served as our key measures. They rated the identity relevance and success relevance of their level of cognitive ability on a sliding scale ranging from 0 (“not at all”) to 100 (“very much”): “To what extent does your level of intuitive intelligence reflect who you are as a person?” and “How important do you think your level of intuitive intelligence is in determining your future success in life?” These measures served as mediator variables, indicative of dissonance-reducing defensive motivation. Finally, as the dependent variable, they rated the subjective importance of their cognitive ability for future happiness on a similar sliding scale (0–100): “How important do you think your level of intuitive intelligence is in determining your future happiness in life?” Finally, all participants were fully debriefed and thanked for their participation.

Results

Identity Relevance. Participants in the poor feedback condition rated their cognitive ability level as less relevant to their personal identity than participants in the good feedback condition ($M = 21.53$, $SD = 22.77$ vs. $M = 62.26$, $SD = 23.32$). The difference was statistically significant and the estimated effect size was very large; $t(101) = 8.97$, $p < .001$, $d = 1.77$).

Success Relevance. Participants in the poor feedback condition also attributed lower importance to cognitive ability level for their future success than participants in the good feedback condition ($M = 28.17$, $SD = 26.58$ vs. $M = 68.00$, $SD = 24.21$). The difference was statistically significant, and the effect size was very large; $t(101) = 7.94$, $p < .001$, $d = 1.56$.

Relevance for Future Happiness. In line with the sour-grape hypothesis, an independent t -test revealed that participants in the poor feedback condition attributed lower importance to their level of cognitive ability for their future happiness than participants in the good feedback condition ($M = 16.06$, $SD = 21.19$ vs. $M = 54.56$, $SD = 26.67$). The difference was statistically significant, and once again, the effect size was very large; $t(101) = 8.14$, $p < .001$, $d = 1.60$.

Mediation Analysis. In line with Hayes's (2013) recommendation, we used a serial mediation model in PROCESS (Model 6) because we had two mediators representing supplementary paths as part of a common theoretical mechanism (cognitive dissonance), rather than representing two competing theories (Model 4). With identity relevance as Mediator 1 and success relevance as Mediator 2, the results were consistent with a model in which the effect of negative test feedback on importance for future happiness is mediated through a sequence of diminished identity relevance and success relevance (see Fig. 3).

Specifically, receiving poor (rather than good) feedback produced lower identity relevance of the cognitive skill ($a_1 = -40.73$, $p < .001$). Lower identity relevance was associated with lower success relevance ($d_{21} = .82$, $p < .001$), which in turn was associated with lower perceived importance of the cognitive ability level for one's future happiness ($b_2 = .51$, $p < .001$). Using 10,000 bias-corrected bootstrapped confidence intervals, this indirect effect ($a_1d_{21}b_2 = -17.21$) was statistically significant, as indicated by a 95% confidence interval not including the value 0 (CI = $-26.24, -8.97$). The total effect of test feedback of importance for future happiness was significant ($c = -38.50$, $p < .001$), while there was no evidence of a direct effect independent of identity relevance and success relevance as mediating mechanisms ($c' = -4.23$, $p = .260$).

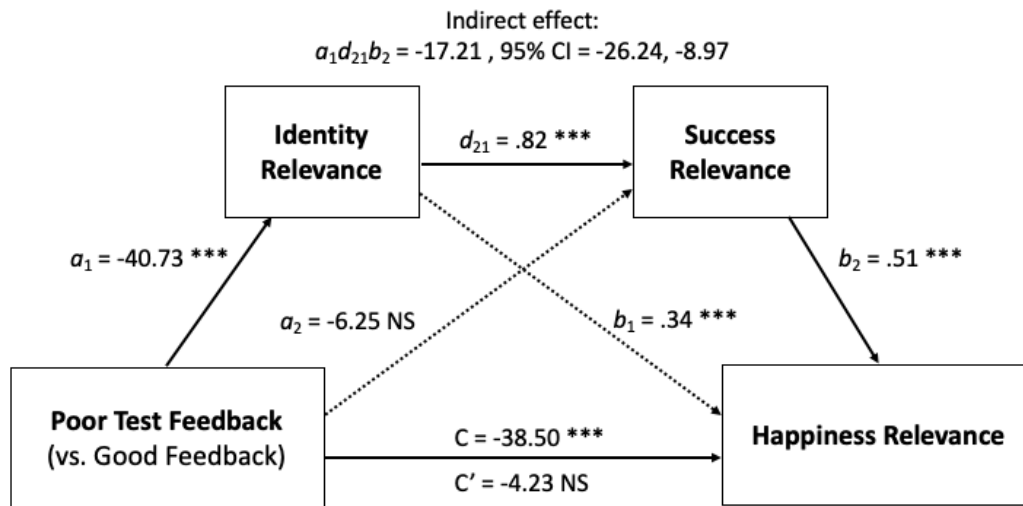


Fig. 3. Statistical mediation diagram for Study 3 (N=103), showing path values as unstandardized regression coefficients. Receiving poor feedback led to perceiving one's performance as less relevant to personal identity and future success than receiving good feedback. Working together in a serial process, reduced identity relevance and reduced success relevance mediated the negative effect of poor test feedback on perceived happiness relevance.

Discussion

Study 3 provided further evidence for the sour-grape hypothesis. Unlike the previous studies, however, Study 3 did not ask people how they would feel about doing better on a similar test in the future. Rather, it asked them to rate how their level of the ability that had been tested would affect their future happiness throughout life. Nonetheless, the results showed that if they did poorly, they thought the ability would not affect their happiness, whereas if they did well, they thought the ability would contribute to their happiness. The differences in terms of standardized effect size were unusually large, suggesting that the sour-grape effect also applies to the attribution of sources or causes of future happiness. While the key question in Study 1 and Study 2 resembled asking someone “Would 10 million dollars make you happy?”, the key question in Study 3 was “Will how much money you have be important for your future happiness?”.

Additional findings illuminated the defensive nature of responses to bad feedback. Participants who were told that their intuitive intelligence was poor thought intuitive intelligence was relatively irrelevant to their self-concept and irrelevant to their ability to

succeed in life. These judgments mediated their conclusion that intuitive intelligence (or, more precisely, their lack of it) would not matter for their long-term happiness. In contrast, people who received positive feedback thought their high intuitive intelligence was important to who they were and would increase their success in life, which in turn would could help them achieve happiness.

Study 4: Grade-Point Salience: A Conceptual Extension

As a test of the generalizability of the sour-grape effect, we designed Study 4 as a conceptual extension without using manipulated variation in performance feedback. Rather, Study 4 capitalized on a naturally occurring variation in initial performance in the form of students' grade-point averages, and merely manipulated the salience of people's grade-point average by asking some participants to report it just *before* predicting how happy they would feel if they were to graduate with an A average. The other half of the sample reported their grade-point average *after* predicting how they would feel.

Participants were recruited among students at a prestigious university program in economics and finance, where grades are close to normally distributed around C+/B- as the average. Less than 5% of the students actually graduate with an A average, making such a lofty goal unlikely (and highly desirable) for most students. We reasoned that asking them to report their grade-point average right before predicting their contingent happiness would make it more salient that the goal was probably unattainable, and that this realization might prompt them to distance themselves from that goal by devaluing it (i.e., by predicting that they would not feel all that happy even if they should reach the goal).

Method

Participants. We recruited 105 participants (50 women; age: $M = 24$ years) from a Norwegian university for an online study of “attitudes and decision-making”. E-mail invitations were sent to students at the Master of Arts level in economics and finance who had one to two years left of their studies before graduating. After completing the experiment, each participant received a lottery ticket in which they could win a gift card at a local coffee shop worth \$5, with a 20% probability of winning.

Materials and Procedure. Participants were asked to respond to three questions presented on separate screens. These questions consisted of reporting their grade-point average at the time, to estimate how likely they thought it was that they would graduate with an average grade of A, and to predict how happy they would feel if they were to graduate with an average grade of A.

The experimental manipulation was the order of these questions. Participants in the salient discrepancy condition were first asked about their grade-point average and the subjective probability of graduating with an A average, and then predicted how happy they would feel if that should happen. The control condition participants predicted their future happiness first and then reported their actual grade-point average and subjective probability of graduating with an A average. Thus, the only systematic difference between the two conditions was whether the negative discrepancy between the ideal and reality was salient prior to the happiness forecast.

Grade-point average was reported on a 15-point scale ranging from A to F (e.g., A, A–, B+, B, B–, C+, etc.). The reported grade was transformed into a number ranging from 1 (A) to 15 (F). Subjective probability was rated on an 11-point response scale ranging from 0 to 10: “Based on your actual grade-point average at the time, how likely do you think it is that you

will graduate from the university with a grade-point average of A?” (0 = “very unlikely,” 10 = “very likely”). The dependent variable measured happiness predictions for an A grade-point scenario: “Imagine that you will graduate from the university with a grade-point average of A. How happy would that make you feel?” (0 = “not at all happy,” 10 = “as happy as you can be”).

At the end of the study, all participants were debriefed and thanked for their participation. All participants were then given true information about the promising future prospects in the job market for students with their kind of education from that particular university, making it clear that in a close to normally distributed grade system, an A average is rare and usually not a prerequisite to getting a good job after graduation.

Results

Descriptive Statistics/Manipulation Check. Across conditions, the self-reported grade-point average was slightly below a B ($M = 4.44$, $SD = 1.85$). As expected, grade-point average did not differ between conditions: Participants who predicted their future happiness before ($M = 4.30$, $SD = 1.91$) versus after ($M = 4.58$, $SD = 1.80$, $t(103) = .760$, $p = .449$) reported similar grades. Only 3 of 105 students reported an A average (2.7%). Across conditions, the subjective probability of graduating with an A average was rated at 2.64 ($SD = 3.22$) on a 0–10 scale. There was no significant difference between conditions in subjective probability for graduating with an A average ($M = 2.81$, $SD = 3.09$ vs. $M = 2.47$, $SD = 3.37$; $t(103) = .60$, $p = .596$). Thus, participants in both conditions evaluated graduating with an A average as highly unlikely, and nearly all students reported a current grade-point average lower than an A.

Predicted Happiness. In line with the sour-grape hypothesis, an independent t -test revealed that participants who had just reported their current grade-point average predicted lower

levels of happiness if they were to graduate with an A average than those who had not yet been reminded of their grades ($M = 6.75$, $SD = 2.14$ vs. $M = 7.70$, $SD = 1.54$). The difference was statistically significant, and the effect size was medium ($t(103) = 2.61$, $p = .010$, $d = .51$).

Thus, being reminded about a salient discrepancy between a lofty future ideal (A) and present reality (B–) made participants predict less happiness following this ideal scenario. Follow-up analyses replicated this finding when controlling for grade-point average, subjective probability estimates, gender, and age as covariates in a univariate analysis of variance ($p = .016$, $d = .49$).

Discussion

Study 4 extended the sour-grape effect in a new and different setting. It used actual performance feedback and relied on individual differences in prior achievement, rather than false feedback. Moreover, it used an established performance metric that is highly relevant and central to the participants' lives (i.e., grade-point average in higher education). It could be argued that in the prior studies, participants could dismiss their test feedback as trivial, but it is difficult for most students to be fully indifferent to their grade-point average at graduation. Most of what they do as students is evaluated by that grade average. Moreover, the manipulation in Study 4 was subtler than that in the preceding studies, which involved taking a test and getting explicit feedback. In Study 4, participants merely answered a seemingly innocent question about their grade-point average either before or after making their happiness forecast, making it a salience manipulation.

Despite these changes, the sour-grape effect obtained once again. In our view, this attests to the generalizability of the effect and to its external validity. Apparently, the defensive devaluation of a goal can be activated merely by being reminded of the distance between one's current achievement level and the idealized (but highly unlikely) goal.

Study 5: Online Experiment—A Combined Replication

To assess the robustness of our central findings in Studies 1–3, we designed Study 5 as a pre-registered replication that combined the key elements from these studies. The logic of design and the main procedures were the same as in Study 1 (relying on a repeated false-feedback paradigm), enabling us to measure both predicted happiness and actual happiness after a second test. We also included the mediator variables from Study 2 and Study 3 (identity relevance and success relevance) to replicate the process mechanism indicative of cognitive dissonance.

There were two main differences between Study 5 and the prior studies. First, Study 5 was incentivized, in that participants learned that they would double their payment if they were to get a top score on the second test. Second, we included established trait measures of individual differences in general need for achievement (McClelland, 1958; Helmreich, Beane, Lucker, & Spence, 1978) and fear of failure (Conroy, Willow, & Metzler, 2002) as moderator variables, to explore potential boundary conditions for the sour-grape effect. We speculated that a high fear of failure would intensify the effects, because fear of failure would increase defensive and self-protective motivations. Meanwhile, we predicted that people with high need for achievement would be less prone than others to protect themselves by withdrawing, because their strong desire to improve would make an initial failure (and the distance to the ambitious goal) less threatening.

The main predictions were that we would replicate the findings from the previous studies. That is, we tested the sour-grape hypothesis, which predicts that initial failure would cause people to devalue the goal by forecasting that a future top performance would not make them very happy. Second, this would be mediated by defensive motivations, and accordingly, participants would rate the ability as not being relevant to their self-concept or to their prospects for success in life. Last, we predicted that this forecasting pattern would turn out to

be wrong, and that actual success would bring ample happiness to participants regardless of their initial score (and regardless of their initial happiness predictions).

Method

Participants. We recruited 200 American participants (128 women; age: $M = 37.5$ years) from an online sample at Amazon Mechanical Turk for a brief study of “personality and cognitive performance” in exchange for \$0.25, with the possibility of doubling their payment to \$0.50 for a top performance.

Materials and Procedure. In this experiment, we pre-registered the hypotheses and statistical analyses in advance of the data collection (PDF available online:

<http://aspredicted.org/sg5va.pdf>).

At the beginning of the study, participants rated themselves on a 19-item need-for-achievement scale (Helmreich et al., 1978) and a 5-item fear-of-failure scale (Conroy et al., 2002) using a 7-point response scale, presented in counterbalanced order. We included these measures as moderator variables. Sample items from the need-for-achievement scale are “I find satisfaction in working as well as I can” and “Part of my enjoyment of doing things is improving my past performance” (1 = “not true at all,” 7 = “exactly true”). Sample items from the fear-of-failure scale are “When I am failing, I am afraid that I might not have enough talent” and “When I am failing, it upsets my “plan” for the future” (1 = “I never think that way,” 4 = “I think that way 50% of the time,” 7 = “I think that way 100% of the time”). Both scales were reliable (fear of failure: Cronbach's $\alpha = .91$; need for achievement: Cronbach's $\alpha = .86$).

Next, participants were provided with two separate Stroop tests consisting of 40 trials each, presented as a reaction-time test of “cognitive speed.” They were told that cognitive speed

is an important part of human intelligence, that it is a result of both talent and effort, and that it is a highly desirable trait that is associated with a broad range of positive outcomes in life. In addition, they were told that a top performance on the main test (i.e., performing in the top 10% of the population) would be rewarded with double payment.

After completing the 40 practice trials, participants received false feedback stating that they either performed poorly (in the bottom 15% of the population) or performed very well (in the top 15% of the population). At this point, all participants answered a couple of questions before moving on to the second test. The mediator variables measured identity relevance and success relevance, as in Study 3. The dependent variable measured how happy participants predicted they would feel if they were to receive a top score on the second test: “Imagine that your level of cognitive speed will be ranked in the top 10% on the actual test. How happy would that make you feel?” (0 = “not at all happy,” 10 = “very happy”). After finishing the second Stroop test, all participants were given manipulated feedback once again, saying that the actual test revealed that their true level of cognitive speed was ranked in the top 10% of the population. As the second dependent variable, they reported their current level of happiness at that point: “In general, how happy do you feel right now?” (0 = “not at all happy,” 10 = “very happy”).

As manipulation checks at the end of study, participants indicated how difficult they thought the first test was (1 = “not at all difficult,” 7 = “very difficult”) and how likely they thought it was to get a top score on the main test after receiving feedback on the practice trial (1 = “very unlikely,” 7 = “very likely”). Last, all participants were fully debriefed and thanked for their participation. All participants were paid the double-payment bonus they “earned” after receiving top score feedback on the second test, to avoid using deceptive economic terms of participation.

Results

Manipulation Checks. Compared with participants in the good feedback condition, participants in the poor feedback condition thought that the first test was more difficult ($M = 4.06$, $SD = 1.54$ vs. $M = 3.24$, $SD = 1.78$; $t(198) = 3.49$, $p < .001$, $d = .49$) and that they would be less likely to get a top score on the second test ($M = 3.73$, $SD = 1.99$ vs. $M = 4.89$, $SD = 1.66$; $t(198) = 4.48$, $p < .001$, $d = .63$). Thus, supporting the internal validity of the experiment, the manipulation successfully altered perceived difficulty and performance expectations of the second test.

Identity Relevance. Those who received poor initial feedback rated their cognitive speed as less relevant for who they were as a person than those who received good feedback ($M = 2.70$, $SD = 2.44$ vs. $M = 5.88$, $SD = 2.85$). The difference was statistically significant, and the estimated effect size was very large; $t(198) = 8.49$, $p < .001$, $d = 1.20$.

Success Relevance. Participants who received poor initial feedback also indicated that their cognitive speed would be less relevant to their future success in life than those who received good feedback ($M = 3.80$, $SD = 2.84$ vs. $M = 6.12$, $SD = 2.54$). The difference was statistically significant, and the effect size was large ($t(198) = 6.08$, $p < .001$, $d = .86$).

Predicted Happiness. Consistent with the sour-grape hypothesis and the previous results, an independent t -test revealed that participants in the poor feedback condition predicted that they would feel less happy after receiving a top score on the second test than participants in the good feedback condition ($M = 7.75$, $SD = 2.21$ vs. $M = 8.39$, $SD = 1.67$). The difference was statistically significant, and the effect size was moderate ($t(198) = 2.31$, $p = .022$, $d = .33$).

Actual Happiness. Also consistent with the previous studies, the level of actual happiness following a top score on the second test was not significantly different between participants in the poor feedback condition and those in the good feedback condition ($M = 8.58$, $SD = 1.69$ vs. $M = 8.81$, $SD = 1.53$; $t(198) = 1.01$, $p = .315$).

In addition to the null effect on actual happiness, we used a difference score (actual happiness – prediction) to assess forecasting accuracy, in which a higher value suggests greater forecasting error. A non-registered follow-up analysis found suggestive evidence that participants in the poor feedback condition underestimated their future happiness to a greater extent than participants in the good feedback condition ($M = 0.83$, $SD = 1.73$ vs. $M = 0.42$, $SD = 1.22$, ($t(198) = 1.94$, $p = .054$, $d = .27$).

Mediation Analysis. As in Study 3, the results from a serial mediation analysis in PROCESS (Model 6) was consistent with a model in which the negative effect of test feedback on happiness forecasting is mediated through a sequence of altered identity relevance (Mediator 1) and success relevance (Mediator 2) (see fig. 4).

Specifically, receiving poor (rather than good) initial feedback reduced identity relevance of cognitive ability ($a_1 = -3.18$, $p < .001$). Reduced identity relevance was associated with reduced success relevance ($d_{21} = .61$, $p < .001$), which in turn was associated with lower happiness forecasts following a top score on the second test ($b_2 = .22$, $p = .001$). Using 10,000 bias-corrected bootstrapped confidence intervals, we showed that this indirect effect ($a_1 d_{21} b_2 = -.42$) was statistically significant, as indicated by a 95% confidence interval not including the value 0 ($CI = -.72, -.19$). Moreover, the total effect of test feedback on happiness forecasting was significant ($c = -.64$, $p = .022$), while there was no evidence of a

direct effect independent of identity relevance and success relevance as mediating mechanisms ($c' = -.33$, $p = .295$).

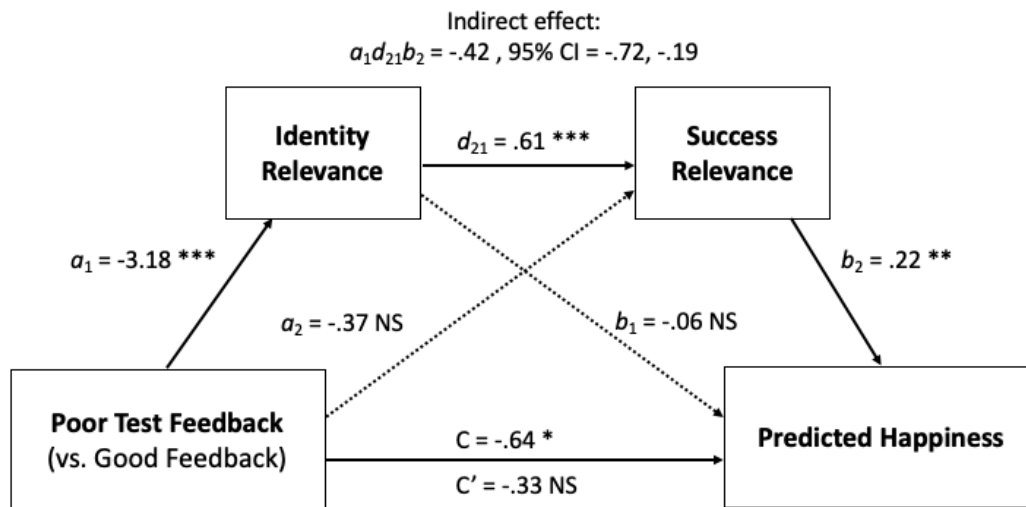


Fig. 4. Statistical mediation diagram for Study 5 ($N=200$), showing path values as unstandardized regression coefficients. Replicating the serial mediation pattern from Study 3, the negative effect of poor test feedback on predicted happiness was mediated through a process of attributing reduced identity relevance and reduced success relevance to one's cognitive performance.

Moderation Analysis: Need for Achievement and Fear of Failure. With happiness

forecasts as the dependent variable, a moderation analysis in the Jamovi software identified a highly significant interaction between experimental condition and need for achievement ($Z = 3.82$, $SE = .27$, $p < .001$). There was no evidence for an interaction effect involving fear of failure ($Z = -.030$, $SE = .19$, $p = .762$). Thus, individual differences in motivation to perform well seemed to influence how participants responded to poor versus good feedback, while individual differences in fear of failing did not.

Therefore, we tested the effect of condition on participants one standard deviation above and one standard deviation below the mean on the need-for-achievement scale, based on 10,000 bias-corrected bootstrap samples (see table 1). These results showed that participants who were low ($-1SD$: $Z = -3.87$, $p < .001$) and average ($Z = -2.17$, $p = .030$) on need for achievement predicted much less happiness after a future top score when they were in the poor feedback condition than when they were in the good feedback condition, whereas participants who were high ($+1SD$: 1.14 , $p < .25$) on need for achievement did not predict less

happiness at all when they were in the poor feedback condition. In other words, we did not find a defensive response to poor initial feedback among people high in achievement motivation, but the sour-grape effect was clearly there for those who had an average or low level of this trait.

Moderation Estimates

	Estimate	SE	Z	p
Condition	-0.580	0.257	-2.26	0.024
Achievement Motivation	0.563	0.138	4.08	< .001
Condition * Achievement Motivation	1.043	0.273	3.82	< .001

Simple Slope Estimates

	Estimate	SE	Z	p
Average	-0.580	0.267	-2.17	0.030
Low (-1SD)	-1.538	0.398	-3.87	< .001
High (+1SD)	0.377	0.331	1.14	0.254

Table 1. Statistical moderation analysis for Study 5 (N=200). With happiness prediction as the dependent variable, the figure illustrates a significant interaction effect ($p = .001$, $Z = 3.82$) of condition (poor vs. good feedback) and individual differences on the need for achievement scale. Unlike participants with low or average achievement motivation, participants with high achievement motivation did not predict less happiness following a future top performance when receiving poor (vs. good) feedback on the practice trial of the test.

Discussion

Study 5 was a pre-registered and incentivized experiment, and provided additional support for the sour-grape effect. This study also added new evidence about mental processes and motivational basis. As in the previous experiments, participants who received poor initial feedback appeared to withdraw defensively. They derogated the ability that they ostensibly lacked (cognitive speed) as being irrelevant to their self-concepts and to their chances for success in life, and they predicted that if they performed superbly on a second test of the same ability they would not experience much happiness from that success. The shift toward viewing the test as irrelevant to one's self-concept and success in life mediated their predictions of lower happiness following a future top score.

As in the preceding studies, however, participants' prediction was wrong: When they took the second test and were told that they had done superbly, they were just as happy as the other participants. One might be tempted to view the mediation patterns as simply rational. After all, does it matter if one scores high on a trait regarded as irrelevant to one's self-concept and life success? But apparently they did care, as indicated by their happiness ratings when they found out that they did actually score well on the test.

We tested two possible trait moderators. One, fear of failure, yielded nothing. The other, achievement motivation, yielded a significant interaction with initial feedback. People high in achievement motivation did not show the sour-grape effect, while the effect was very strong for people with low achievement motivation. Taken together, these findings offer insight into what creates the boundaries for defensive devaluation of future success. It is apparently a high desire to succeed, rather than a lack of fear of failure, that enables people to avoid distancing themselves from future success. We could speculate that failure is at least somewhat disturbing to everyone, so individual differences in fear of failure did not moderate the sour-grape response. Meanwhile, though, high need for achievement can be sufficiently compelling that one continues to view success as highly desirable even when it seems improbable and (for the moment) out of reach.

Study 6: Online Experiment—A High-Powered, Extended Replication

As our final assessment of the sour-grape effect, Study 6 was a pre-registered and high-powered replication. Inspired by the "explore small-confirm big" strategy (Sakaluk, 2016) and the "small telescopes" approach to ensure robust and replicable research (Simonsohn, 2015), we recruited a sample that was more than three times larger than Study 5. This enables detection or rejection of a much smaller effect size than what we have observed previously, and allows for a more precise estimation of the true effect.

Making this an extended replication, we used a broader outcome measure of emotional well-being than the previous experiments. Whereas the previous studies used a single-item measure of predicted happiness, Study 6 used a three-item measure that included predicted happiness, predicted pride, and predicted gratitude. The hypothesis was that people who receive poor (vs. good) test feedback would predict lower positive emotion following a future top performance. We included no moderator or mediator variables in this study.

Method

Participants. We recruited 650 American participants from an online sample at Amazon Mechanical Turk for a brief study of “personality and response-time” in exchange for \$0.50. After excluding 36 individuals who failed a simple attention check at the beginning of the survey, the final sample consisted of 614 participants (324 women; age: $M = 39$ years).

According to a sensitivity analysis in the G*power software, the final sample provided 90% power to detect a true effect of $d = .26$ or larger ($p < .05$, two-tailed). In our calculation, the average effect size of test feedback on happiness forecasting in the previous experiments (studies 1-5) is $d = .58$. This suggest that this final replication had 90% power to detect an effect that is 45% as large as the effect previously observed, and in consequence, very high power to detect a larger effect in the similar range as before.

Materials and Procedure. We pre-registered the hypothesis and statistical analysis of Study 6 in advance of the data collection (PDF available online:

<https://aspredicted.org/jh6im.pdf>).

After reporting age and gender, the participants read an identical description of the cognitive trait in question ("cognitive speed") as in Study 5. Then, unlike the previous experiments, the participants rated how desirable they thought it was to perform well on the

given task: "What is your ideal level of cognitive speed? Specifically, what percentage of the population would you like to outperform on established measures of cognitive speed?"

(sliding scale: Better than 0% - Better than 100%). This measure enabled an assessment of whether the participant sample started out appreciating the importance of the cognitive trait, and whether they actually wanted to perform well on the test.

In the main part of the study, the participants completed a 50-trial Stoop test, followed by false feedback stating that they either performed poorly (bottom 15% of the population) or performed very well (top 15% of the population). The dependent variable, predicted positive emotion, was measured with three items on the next screen. 1) Predicted happiness: "Imagine that one month from now, you take a new test and learn that your actual level of cognitive speed is ranked in the top 10% of the population. How happy would that make you feel?" (0 = "not at all happy," 10 = "as happy as you can be"). 2) Predicted pride: "How proud would that make you feel?" 0 = "not at all proud," 10 = "as proud as you can be"). 3) Predicted gratitude: "How grateful would that make you feel?" 0 = "not at all grateful," 10 = "as grateful as you can be"). The responses on the three items were highly correlated, suggesting that the measure was a reliable assessment of predicted positive emotion (Cronbach's alpha = .95).

As a simple manipulation check at the end of study, participants reported how they would describe the feedback they received on the cognitive test (1 = positive, 2 = neutral, 3 = negative). Last, as in the previous experiments, all participants were fully debriefed and thanked for their participation.

Results

Descriptive Statistics and Manipulation Check.

At the beginning of the experiment, the participants rated it as a highly desirable to perform well on the cognitive test. The average response was an aspiration of outperforming 72.5% of

the population on a test of "cognitive speed" ($SD = 18.5$). As expected, there was no notable difference between the two conditions prior to the test (73% vs 72%, $p < .54$). Thus, the participants started out with an equal desire of performing well.

The manipulation check indicated that participants correctly perceived their own feedback. The scale ran from 1 (good) to 3 (bad), and the ratings in the negative feedback condition were significantly worse ($M = 2.67$, $SD = .60$) than in the positive feedback condition ($M = 1.07$, $SD = .25$; $t(612) = 42.92$, $p < .001$, $d = 3.48$). Thus, the experimental manipulation successfully altered their perceived performance on the cognitive test.

Hedonic prediction: Happiness, pride and gratitude. Consistent with the pre-registered hypothesis and replicating previous results, an independent t -test found that participants in the poor feedback condition predicted that they would feel less positive emotion after a future top score than participants in the good feedback condition ($M = 5.90$, $SD = 3.22$ vs. $M = 7.64$, $SD = 1.92$). The difference was statistically significant, and the estimated effect size was above medium in strength ($t(612) = 8.08$, $p < .001$, $d = .65$).

In a non-registered follow-up analysis, we repeated this analysis for each specific emotion prediction, and found a similar effect on all three (fig. 5). Specifically, participants who received poor (vs. good) feedback on the cognitive test predicted less happiness ($M = 6.15$, $SD = 3.29$ vs. $M = 7.85$, $SD = 1.82$, $p < .001$, $d = .64$), less pride ($M = 5.95$, $SD = 3.40$ vs. $M = 7.83$, $SD = 2.05$, $p < .001$, $d = .67$), and less gratitude ($M = 5.60$, $SD = 3.37$ vs. $M = 7.22$, $SD = 2.32$, $p < .001$, $d = .56$) if they should achieve a top result in the future.



Fig. 5. Study 6 (N=614): In line with the sour-grape hypothesis, participants receiving poor (vs. good) feedback on a cognitive test predicted lower levels of positive emotion following a future top performance (happiness: $p < .001$, $d = .64$; pride: $p < .001$, $d = .67$; gratitude: $p < .001$, $d = .56$). Error bars indicate standard error. All three outcome variables were measured using a 0–10 response scale.

Discussion

Study 6 provided a high-powered and extended replication of the sour-grape effect. Despite the fact that participants in both conditions started out with a high desire to perform well on the test, the participants who received poor feedback predicted less positive emotion following a future top performance than participants who received good feedback.

In our view, these results further support a theoretical explanation emphasizing defensive motivation to protect the self: The participants who received poor feedback and then predicted low positive emotion did indeed *revise* and reduce their judgment of how desirable it would be for them to succeed in the future. This was true both when using a broad index measure of hedonic prediction consisting of happiness, pride and gratitude, and when analyzing each predicted emotion individually.

Internal Meta-Analysis

As a statistical summary of the sour-grape effect on happiness prediction, we conducted an internal meta-analysis (e.g., Goh, Hall, & Rosenthal, 2016) across all six experiments in this

investigation. Using a random effects model in the JASP software (2020), the basis for this analysis was the standardized effect size (Cohen's d) and the standard error for each effect.

Predicted Happiness

By including one effect size from each of the six studies ($k = 6$, total $N = 1,304$), the average effect of experimental condition on happiness prediction was $d = 0.66$, indicative of an above medium effect size ($Z = 3.76$, $p < .001$, two-tailed). Thus, in line with the sour-grape hypothesis, people predicted less future happiness when poor feedback or a performance reminder made the goal seem remote or unattainable (see figure 6).

In addition, Studies 1 and 5 enabled us to assess forecasting accuracy, by comparing happiness predictions with the actual level of experienced happiness after a second test. Across these two studies ($k = 2$, total $N = 282$), the average effect of experimental condition on happiness underestimation was $d = 0.34$, indicative of a moderate effect size ($Z = 2.87$, $p = .004$, two-tailed). Thus, participants in the poor feedback condition underestimated their future happiness to a larger extent than participants in the good feedback condition.

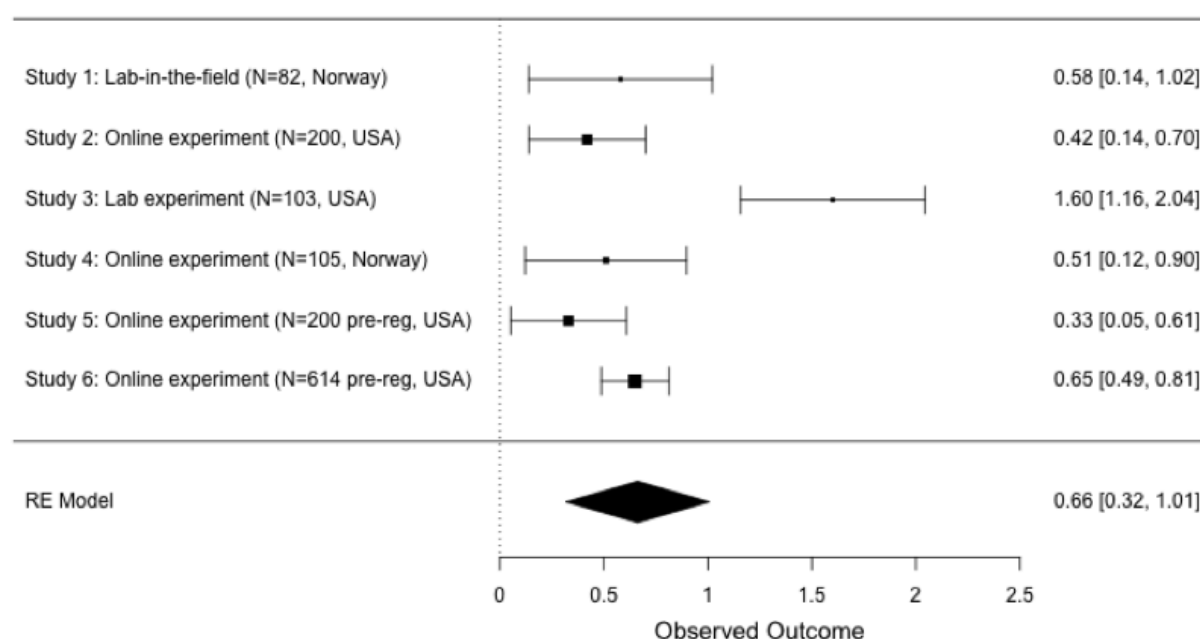


Figure 6. Internal meta-analysis of happiness predictions across all six studies (N=1,304). Participants receiving poor (vs. good) feedback on their cognitive performance predicted that they would feel less happy following a future top score. Predicted happiness was measured on a 0–10 scale (Studies 1, 4–6) or a 0–100 scale (Studies 2 and 3¹). All experiments used a single-item measure of happiness, except Study 6 which used a 3-item measure (happiness+pride+gratitude/3). Error bars indicate 95% confidence intervals, and is stated in the parentheses following each effect size. The meta-analytic effect size is located in the bottom right corner ($d = 0.66, p < .001$).

General Discussion

With a total sample of 1,304 participants across two different countries and two different languages, we used six experiments to examine how people deal with failure when evaluating future goals. The results provided strong support for the sour-grape hypothesis: People predicted that their happiness over a future success would be much lower if their initial performance had been poor rather than good, but when they succeeded on a second test they actually became equally happy as everyone else. That is, initial failure made people devalue and underestimate how good it would feel to succeed in the future.

Even when student participants were simply reminded of their current grade-point average and the poor chances of graduating with a perfect A average, they downgraded how

¹ The outcome measure in Study 3 was the *perceived relevance* of the cognitive performance for one's future happiness in life, rather than a pure happiness prediction as in the other studies. We included this modified measure in Study 3 to enable a broader test of the sour-grape hypothesis.

happy they thought they would be if that should happen. In a final replication, the sour-grape effect was found once again and generalized from happiness predictions to a broader measure of positive emotion, including predictions of pride and gratitude.

We note that this finding is far from obvious. Both scarcity theories from psychology and the law of supply-and-demand from economics suggest that the opposite effect, the "greener-grass hypothesis", would be supported: As a desirable outcome becomes less accessible, its subjective value should increase. Indeed, we surveyed 100 individuals at Amazon's Mechanical Turk (reported in the Appendix), and asked them to predict the outcome of our first study. Perfectly in line with the greener-grass hypothesis, and in stark contrast to our actual findings, they predicted that participants improving from a poor test result to a superb one (i.e., going from 2/6 correct to 6/6 correct) would feel significantly happier than those who attained a similar top score at the end but performed quite well from the start (i.e., going from 5/6 correct to 6/6 correct). Thus, when people viewed the current test situation from a distance, without experiencing the pain of failure themselves, they failed to anticipate the defensive response of our participants.

Sour Grapes: Interpretation of Mechanisms

Various findings fit the view that people downgrade their hedonic forecasts as a way to protect themselves from possible disappointment (e.g., Shepperd et al., 2000). In the current set of studies, the downgrading of future happiness for potential success was mediated by asserting that the ability being tested was not relevant to one's self-concept and probably not relevant for one's success in life either. Although our investigation was inspired in part by Aesop's fable of the fox, there are some important differences. The fox only derogated the grapes after he was sure he could not get them, whereas for participants in our studies, success remained possible. This is also a notable difference from previous research focusing on

adjustment to uncontrollable events (e.g., Wilson, Wheatley, Kurtz, Dunn, & Gilbert, 2004) and learning from previous experience (Wilson, Meyers, & Gilbert, 2001). More important, in terms of process mechanisms, the fox did not arrive at the conclusion that the grapes were sour by way of asserting that they were irrelevant to his self-concept or that being able to eat grapes in general was irrelevant to his success in life. But that was what we found.

Whether the fox lacked achievement motivation was also left unspecified by Aesop and previous research, but Study 5 in the present investigation showed that people with high achievement motivation were less likely than others to exhibit the sour-grape effect. That is, people who were strongly motivated to achieve success in general did not dismiss their goal after an initial, discouraging failure. Possibly reflecting their intrinsic desire to learn and improve through trial and error, they continued to believe that they would be quite happy if they were to succeed, and they were correct. This finding seems consistent with previous research on the process of goal pursuit (Fishbach, Koo, & Finkelstein, 2014), in which individuals with strong (vs. weak) commitment become more motivated by attending to the *lack* of accomplishment, experiencing greater desire to make progress towards the goal.

The present investigation yields converging evidence for the sour-grape effect. Indeed, the pattern of findings was highly consistent despite changes in methods and samples. The studies included an experiment in a field setting, online experiments, and a laboratory experiment; they used both community, online and student populations; they were run on two different continents in two different languages; and they included two pre-registered replications. Convergent evidence across diverse methods and samples increase confidence in the generality and validity of the conclusions.

Limitations

Despite the contextual and methodological variation between our studies, they were all conducted in western, educated, industrialized, rich and democratic countries (USA and Norway). Thus, cross-cultural generalizations of the present findings should be done with caution, and ideally, future research will explore generalizability to culturally very different samples (Henrich, Heine, & Norenzayan, 2010; Tiokhin, Hackman, & Hruschka, 2018).

As a second limitation, we note that the mediation analyses reported in this investigation cannot confirm that our conceptual model is the only plausible alternative. In our view, strong causal inference is only permitted for the direct effects between the manipulated independent variable and each of the observed outcome variables, whereas the correlational relationship between the non-manipulated outcome variables is merely suggestive for the underlying process. We therefore refer to the mediation results as *consistent with* our proposed model of the causal process (Hayes, 2013), in which the negative effect of poor test feedback on happiness prediction is driven by reduced evaluations of personal and future relevance.

Third, we acknowledge that the two-condition design in our experiments, comparing the effect of good versus bad feedback, could have benefitted from a no-feedback control as a third condition. Although this would result in lower statistical power and lower test precision (given the same sample size), such a design could provide additional information about the relative effect of good versus bad feedback as compared to a "neutral" control. For instance, it could be the case that poor feedback leads to reduced valuation of future goals, whereas good feedback leads to *higher* valuation than no feedback. This and other possibilities remain to be examined in future research. Nevertheless, it seems reasonable to assume that the present findings fit the widely found pattern in which bad or negative events have stronger effects than good or positive ones (see Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001).

Finally, in terms of language framing, we note that the future goal in some of our experiments was construed as rather unlikely (e.g., Study 1: "If you *manage* to get 6/6 correct on the second test"), or preceded by a question where participants had to reflect about the low likelihood of a future success (Study 4). Future research should examine whether the specific wordings used in the current investigation is a boundary condition for the sour-grape effect to emerge, or whether the same effect will generalize across different descriptions of the goal and different specifications of well-being (e.g., happiness with the final outcome vs. happiness in general). For replication studies or extensions of the current work, we recommend our Study 6 as a natural choice, given the high statistical power and robust main effect in this pre-registered experiment.

Implications

The current findings constitute an exception to the well-documented impact bias pattern in affective forecasting (for review, see Morewedge, 2014). The typical pattern is that people overestimate the intensity and duration of future feelings. Our findings show a sour-grape effect in the opposite direction: Underestimation of future happiness as a consequence of initial failure. Wilson and Gilbert (2005) speculated that some exceptions might arise because making predictions in a "cold" and rational state might fail to appreciate "hot" motivational reactions in the future (see Loewenstein et al., 2003; Loewenstein, 2005). Attempting to reconcile previous findings, a recent study showed that differences in outcome specification can predict the direction of forecasting errors (Buechel, Zhang, & Morewedge, 2017).

Our work suggests another category of exception, informed by cognitive dissonance theory (Festinger, 1957; Aronson, 1992) and the principle of "adaptive preferences" (Elster, 1983), namely when the person is defensively motivated to deny the emotional relevance of

future achievements. The defensive nature of this hypothesis is indicated by multiple findings, starting with the discrepancy between people's predictions about how they would feel and how they actually felt when the unlikely success was achieved. The mediation pattern also attests to defensive motivation: When people thought they had performed poorly on the practice trial of a cognitive test, they stated that their level of cognitive ability was not relevant for their personal identity or future success in life. However, if that was the case, one should expect that these participants would care less about their next test result and be less happy after receiving a top score than other participants. As we now know, this did not occur.

This finding provides new insight into the flexibility of self-protective processes: People seem to naturally identify with the things they are good at, either in present reality or their future expectations. Despite the negative predictions made in advance, the pain of past failure was not enough to reduce the pleasure of a new success, once the positive outcome was actually achieved.

The motivated underestimation of future happiness could fit well with the broader theoretical framework. Various writers have suggested that impact bias is adaptive because overestimating future emotions can motivate people to work harder toward achievable goals (e.g., Norem & Cantor, 1986; Baumeister, Vohs, DeWall, & Zhang, 2007; Morewedge & Buechel, 2013; Miloyan & Suddendorf, 2015). However, disengaging from unattainable goals is also useful because it prevents people from wasting their time and energy in futile pursuits (e.g., Henderson, Gollwitzer, & Oettingen, 2007). Self-regulation research has pointed out that remaining attached to unrealistic fantasies for the future may actually harm motivation and goal progress (Oettingen, 2012), and especially when potential obstacles and strategic plans are not considered (Gollwitzer & Sheeran, 2006). Although a discrepancy between the present situation and the ideal state can be motivating in some cases and for some individuals (Koo & Fishbach, 2014; Fishbach, Koo, & Finkelstein, 2014), our findings suggest that

people respond with self-protective disengagement when the distance to the goal becomes too big. Some evidence shows that disengaging from unattainable goals and shifting efforts to different, more attainable goals enhances well-being (Wrosch, Scheier, Miller, Schulz, & Carver, 2003). The sour-grape effect may be a helpful part of this adaptive disengagement process, insofar as doing so redirects one's attention to more achievable goals.

Concluding Remarks

Drawing on Festinger's (1957) theory of cognitive dissonance, the philosopher and decision theorist Jon Elster (1983) defined an adaptive preference as a desire that has adapted to what is possible. The idea of humans as perfectly rational and utility-maximizing agents is widespread in traditional economics and other normative decision theories. Challenging those views, the principle of adaptive preferences implies that people "satisfice" (Simon, 1956) their wants based on (easily) available possibilities. That is, people rarely aim for the maximally best outcome when making choices and pursuing goals — instead, they select the first alternative in the choice set that appears "good enough" and within reach.

The sour-grape effect provides further evidence of motivated bias, and yet this effect may be part of a broader pattern in which people adjust their motivational and emotional responses in light of the shifting perceptions of what is possible (Simon, 1956; Elster, 1983). We speculate that such a pattern is indeed adaptive and useful, as a general heuristic in everyday decision-making. Devaluing unattainable goals might help the person to maintain a positive self-concept and allocate limited resources elsewhere. A possible exception from such adaptive functions would arise if people misperceive attainable goals as unattainable, for example, due to misattributing the reason for failure or underestimating the effectiveness of hard work. Such and other possibilities regarding the function of the sour-grape effect remain for further research.

Folk wisdom is notorious for offering contradictory prescriptions. Yet in the case of how people deal with personal failure, we found consistent evidence for the sour-grape effect, following Aesop's lead. Again and again, believing that a future success was out of reach made people dismiss how good it would feel to achieve it.

Appendix: Prediction Survey

The six experiments in the current investigation, are concerned with how people predict their own happiness with respect to a seemingly likely versus unlikely success. As an example, in Study 1, the participants received false feedback on a first test of their "intuitive intelligence", which was either presented as poor (2/6 correct) or quite good (5/6 correct). After receiving this feedback, the participants predicted how happy they would feel if they should get a top score on the second test (6/6 correct). Finally, after all participants received such a top score, they reported how happy they actually felt on the same scale. Across all our six studies, we found strong support for the "sour-grape effect": The participants receiving poor (vs. good) feedback predicted to feel *less* happy after a future top performance, but when they received such a top score they ended up being just as happy as anyone else (and happier than they predicted in advance).

However, in this survey, we asked a group of 100 participants to predict how *other people* would respond in the specific experimental situation. The purpose with this survey was to compare people's lay beliefs with how people actually responded to initial failure.

Method

Participants. We recruited 100 American participants (55 women; age: $M = 36.5$ years) from an online sample at Amazon Mechanical Turk for a brief study of "attitudes and social judgment" in exchange for \$0.25.

Materials and Procedure. Based on the experimental procedure from Study 1, all participants were asked to read a text consisting of two scenarios leading up to a top score on a cognitive test (one condition, within-subject). The task for the participant was to rate which one of the two people involved would be happiest about the final outcome. The only difference noted between the two individuals was that person A performed poorly on a practice trial of the test while person B performed very well. After this difference in initial test performance, they both performed as well as they possibly could on the actual test, getting a top score on the test and winning a bonus. The text of the two scenarios read as follows:

Imagine that person A and person B are participating in a research study on “intuitive intelligence,” which is a cognitive ability that is strongly related to general intelligence and positive outcomes in life. The best possible score on this test is getting 6 out of 6 correct, which is indicative of excellent cognitive ability. Before the actual test, person A and person B conducted a practice trial where they received feedback on their performance.

Imagine that person A scored only 2/6 correct on the first test while person B scored 5/6 correct. However, on the actual test conducted next, both of them scored a perfect 6/6, indicative of excellent cognitive ability.

After reading this text, all participants rated the relative happiness of person A and person B, which was the dependent variable in this survey: “If there should be any difference in how happy person A and person B felt after receiving a top score on the second test, who of them do you think would feel happier about it?” (1 = “clearly person A,” 4 = “no difference,” 7 = “clearly person B”).

Results

Social Judgment of Relative Happiness. We ran a one-sample *t*-test to determine respondents’ intuitions about which target person would feel happier about achieving a top score on the final test. The relative mean rating favored person A over person B ($M = 2.51$, $SD = 1.90$), and the mean difference was 1.49 below the “no difference” value of 4 on the

response scale (95% confidence interval [CI]: $-1.87, -1.11$). This difference was statistically significant ($t(99) = -7.83, p < .001$).

Thus, participants thought that happiness would be higher for someone who scored a perfect 6 after getting only 2 right on the practice round than for someone who scored 6 after getting a 5 on the practice round. As such, lay intuition on the behalf of others appear to favor the greener-grass hypothesis: Unexpected success brings more joy than expected success.

Discussion

In the paper, a series of six experiments found a pattern in self-predicted happiness that support the sour-grape hypothesis. Participants who were randomly assigned to receive poor feedback on the first test predicted lower happiness in reaction to a future top score than participants who had been almost perfect the first time.

The prediction survey reported here, focused on social judgments of how others would respond in a similar situation as in Study 1, and yielded the opposite result. Apparently, when predicting the happiness of other people, lay intuition favors the greener-grass hypothesis, while the sour-grape hypothesis fits people's actual predictions on the behalf of themselves. More precisely, people thought others would be happier when improving from 2 to 6 correct than when improving from 5 to 6 — but when people took the tests themselves, they predicted less happiness in the 2-to-6 condition than in the 5-to-6 condition.

Part of the reason for the difficulty in predicting this result, might be that judgments about oneself are driven by defensive motivation and self-protective concerns to a greater extent than judgments about others (e.g., Gilovich et al., 2002; Kahneman, 2011), suggesting a cold-hot empathy gap between forecasts from the "outside view" and forecasts from the "inside view" by people who actually experienced failure themselves (Loewenstein, 2003;

2005). In all cases, we conclude that the sour-grape effect is a robust effect in hedonic prediction, although it is far from obvious for the outside observer.

Open Science Statement:

As internal replication, the primary hypotheses and statistical analyses in Study 5 and 6 were pre-registered in advance of each data collection. Data and materials for all six studies, and pre-registrations for Study 5 and 6, are openly available:

https://osf.io/tsdx4/?view_only=8a8b4b0cdf514c96937c56ee5cdb96d1.

Acknowledgements:

This work was partly supported by the Research Council of Norway through its Centres of Excellence Scheme, FAIR project no 262675.

We thank Annika Rødeseike and Ingrid Toppe for assistance with the data collection for Study 1, and Siv Skard for assistance with the statistical analyses. We also thank Helge Thorbjørnsen, Einar Breivik, Sigurd Troye, Denise de Ridder and George Loewenstein for providing useful feedback on earlier versions of the manuscript, and the editor (Marlone Henderson) and two anonymous reviewers for securing a constructive review process that helped refine and improve the final paper.

Author contact:

Correspondence concerning this article should be addressed to Hallgeir Sjøstad, FAIR Insight Team, Centre for Applied Research at NHH and Norwegian School of Economics.

E-mail: Hallgeir.Sjastad@snf.no

References

- Aronson, E. (1992). The return of the repressed: Dissonance theory makes a comeback. *Psychological Inquiry*, 3(4), 303–311.
- Arrowood, A. J., & Ross, L. (1966). Anticipated effort and subjective probability. *Journal of Personality and Social Psychology*, 4(1), 57.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5, 323–370.
- Baumeister, R. F., Maranges, H. M., & Sjøstad, H. (2018). Consciousness of the future as a matrix of maybe: Pragmatic prospection and the simulation of multiple possibilities. *Psychology of Consciousness: Theory, Research, and Practice*, 5(3), 223.
- Baumeister, R. F., Vohs, K. D., DeWall, C. N., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, 11(2), 167–203.
- Bentham, J. (1789). *The principles of morals and legislation*. New York: Prometheus Books.
- Berglas, S., & Jones, E. E. (1978). Drug choice as a self-handicapping strategy in response to noncontingent success. *Journal of Personality and Social Psychology*, 36(4), 405.
- Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology*, 36(8), 917.
- Bruni, L., & Sugden, R. (2007). The road not taken: how psychology was removed from economics, and how it might be brought back. *The Economic Journal*, 117(516), 146–173.
- Buechel, E. C., Zhang, J., & Morewedge, C. K. (2017). Impact bias or underestimation? Outcome specifications predict the direction of affective forecasting errors. *Journal of Experimental Psychology: General*, 146(5), 746–761.
- Cialdini, R. (1993). *The psychology of influence*. New York: William Morrow & Co.
- Conroy, D. E., Willow, J. P., & Metzler, J. N. (2002). Multidimensional fear of failure measurement: The performance failure appraisal inventory. *Journal of Applied Sport Psychology*, 14(2), 76–90.
- DeSteno, D. (2018). *Emotional success: The power of gratitude, compassion, and pride*. Houghton Mifflin Harcourt.
- DeWall, C. N., Baumeister, R. F., Chester, D. S., & Bushman, B. J. (2016). How Often Does Currently Felt Emotion Predict Social Behavior and Judgment? A Meta-Analytic Test of Two Theories. *Emotion Review*, 8(2), 136–143.
- Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge University Press (CUP).
- Ent, M. R., & Gerend, M. A. (2016). Cognitive dissonance and attitudes toward unpleasant medical screenings. *Journal of health psychology*, 21(9), 2075–2084.
- Fein, D., Gleeson, M. K., Bullard, S., Mapou, R., & Kaplan, E. (1998). *The Biber Cognitive Estimation Test*. Poster presented at the annual meeting of the International Neuropsychological Society, Honolulu, HI.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.
- Fishbach, A., Koo, M., & Finkelstein, S. R. (2014). Motivation resulting from completed and missing actions. In *Advances in experimental social psychology* (Vol. 50, pp. 257–307). Academic Press.

- 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(3), 617.
- Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta-Analysis of Your Own Studies: Some Arguments on Why and a Primer on How: Mini Meta-Analysis. *Social and Personality Psychology Compass*, 10(10), 535–549.
- Golub, S. A., Gilbert, D. T., & Wilson, T. D. (2009). Anticipating one's troubles: The costs and benefits of negative expectations. *Emotion*, 9(2), 277.
- Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in experimental social psychology*, 38, 69-119.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.
- Helmreich, R. L., Beane, W., Lucker, G. W., & Spence, J. T. (1978). Achievement motivation and scientific attainment. *Personality and Social Psychology Bulletin*, 4(2), 222–226.
- Henderson, M. D., Gollwitzer, P. M., & Oettingen, G. (2007). Implementation intentions and disengagement from a failing course of action. *Journal of Behavioral Decision Making*, 20(1), 81–102.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33 (2-3), 61-83.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Kahneman, D., & Snell, J. (1990). Predicting utility. In Hogarth, R. M. (Ed). *Insights in decision making: A tribute to Hillel J. Einhorn (pp. 295-310)*. Chicago, IL: University of Chicago Press.
- Kahneman, D., & Snell, J. (1992). Predicting a changing taste: Do people know what they will like? *Journal of Behavioral Decision Making*, 5(3), 187–200.
- Kay, A. C., Jimenez, M. C., & Jost, J. T. (2002). Sour grapes, sweet lemons, and the anticipatory rationalization of the status quo. *Personality and Social Psychology Bulletin*, 28(9), 1300–1312.
- Lepper, M. R., Zanna, M. P., & Abelson, R. P. (1970). Cognitive Irreversibility in a dissonance-reduction situation. *Journal of Personality and Social Psychology*, 16(2), 191.
- Loewenstein, G. (2005). Hot-cold empathy gaps and medical decision making. *Health Psychology*, 24(4S), S49.
- Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics*, 118(4), 1209–1248.
- McClelland, D. C. (1958). Methods of measuring human motivation. *Motives in Fantasy, Action, and Society*, 7–42.
- Mellers, B. A., & McGraw, A. P. (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science*, 10(6), 210–214.
- Miloyan, B., & Suddendorf, T. (2015). Feelings of the future. *Trends in Cognitive Sciences*, 19(4), 196–200.
- Morewedge, C. K., & Buechel, E. C. (2013). Motivated underpinnings of the impact bias in affective forecasts. *Emotion*, 13(6), 1023–1029.

- Morewedge, C. K. (2014). *Utility: Anticipated, experienced, and remembered*. In G. Keren & G. Wu (Eds.). *Blackwell Handbook of Judgment and Decision Making*, 2nd Edition. Malden, MA: Blackwell Press.
- Norem, J. K., & Cantor, N. (1986). Defensive pessimism: harnessing anxiety as motivation. *Journal of Personality and Social Psychology*, 51(6), 1208.
- Oettingen, G. (2012). Future thought and behaviour change. *European review of social psychology*, 23(1), 1-63.
- Read, D. (2007). Experienced utility: utility theory from Jeremy Bentham to Daniel Kahneman. *Thinking & Reasoning*, 13(1), 45–61.
- Roberts, W. A. (2002). Are animals stuck in time? *Psychological Bulletin*, 128(3), 473.
- Sakaluk, J. K. (2016). Exploring small, confirming big: An alternative system to the new statistics for advancing cumulative and replicable psychological research. *Journal of Experimental Social Psychology*, 66, 47–54.
- Shaffer, D. R., & Hendrick, C. (1971). Effects of actual effort and anticipated effort on task enhancement. *Journal of Experimental Social Psychology*, 7(4), 435-447.
- Shallice, T., & Evans, M. E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, 14(2), 294–303.
- Shepperd, J. A., Findley-Klein, C., Kwavnick, K. D., Walker, D., & Perez, S. (2000). Bracing for loss. *Journal of Personality and Social Psychology*, 78(4), 620–634.
- Shepperd, J. A., & McNulty, J. K. (2002). The affective consequences of expected and unexpected outcomes. *Psychological Science*, 13(1), 85–88.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological science*, 26(5), 559-569.
- Suddendorf, T., & Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs*, 123(2), 133–167.
- Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, 30(03), 299–313.
- Sweeny, K., & Shepperd, J. A. (2010). The costs of optimism and the benefits of pessimism. *Emotion*, 10(5), 750.
- Temple, R., Temple, R. K. G., & Temple, R. (1998). *The complete fables by Aesop*. Penguin Classics, New York.
- Tiokhin, L., Hackman, J., & Hruschka, D. (2018, January 17). Why replication is not enough: Insights from a cross-cultural study of social discounting. Working paper. Retrieved from psyarxiv.com/f5s84
- Wilson, T. D., Wheatley, T. P., Kurtz, J. L., Dunn, E. W., & Gilbert, D. T. (2004). When to fire: Anticipatory versus postevent reconstrual of uncontrollable events. *Personality and Social Psychology Bulletin*, 30(3), 340-351.
- Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting knowing what to want. *Current Directions in Psychological Science*, 14(3), 131–134.
- Wilson, T. D., Meyers, J., & Gilbert, D. T. (2001). Lessons from the past: Do people learn from experience that emotional reactions are short-lived?. *Personality and Social Psychology Bulletin*, 27(12), 1648-1661.

HOW PEOPLE VALUE FUTURE GOALS AFTER INITIAL FAILURE

Wrosch, C., Scheier, M. F., Miller, G. E., Schulz, R., & Carver, C. S. (2003). Adaptive self-regulation of unattainable goals: Goal disengagement, goal reengagement, and subjective well-being. *Personality and Social Psychology Bulletin*, 29(12), 1494–1508.