Research Article

Not Learning From Failure—the Greatest Failure of All

Lauren Eskreis-Winkler and Ayelet Fishbach
Booth School of Business, University of Chicago

Abstract
Our society celebrates failure as a teachable moment. Yet in five studies (total $N = 1,674$), failure did the opposite: It undermined learning. Across studies, participants answered binary-choice questions, following which they were told they answered correctly (success feedback) or incorrectly (failure feedback). Both types of feedback conveyed the correct answer, because there were only two answer choices. However, on a follow-up test, participants learned less from failure feedback than from success feedback. This effect was replicated across professional, linguistic, and social domains—even when learning from failure was less cognitively taxing than learning from success and even when learning was incentivized. Participants who received failure feedback also remembered fewer of their answer choices.

Why does failure undermine learning? Failure is ego threatening, which causes people to tune out. Participants learned less from personal failure than from personal success, yet they learned just as much from other people’s failure as from others’ success. Thus, when ego concerns are muted, people tune in and learn from failure.

Keywords
learning, feedback, failure, ego threat, motivation, open data, open materials, preregistered

Received 5/16/19; Revision accepted 9/5/19

Our society celebrates failure as a teachable moment (Brown, 2014; Maxwell, 2007). Research appears to support this point. People react more strongly—physiologically, cognitively, and emotionally—to negative events than positive ones (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Kahneman & Tversky, 1979; Rozin & Royzman, 2001; Taylor, 1991) in ways that arguably enhance learning. For example, compared with positive stimuli, negative stimuli command more attention (Öhman, 2007; Pratto & John, 1991) and increase information processing (Bless & Fiedler, 2006; Ohira, Winton, & Oyama, 1998; Puig & Szpunar, 2017; Taylor, 1991). It follows that people may pay attention to failure, process it, remember it, and thus learn from it—as much or more than they learn from success.

Here, we explored the alternative: that people learn less from failure. Whether people learn from failure depends not only on whether failure is attention grabbing but also on people’s motivation to attend to it. If people are motivated to ignore their failures, then they will not attend to them and will not learn from them. For example, if researchers are motivated to ignore failed experiments, they will learn nothing from them.

Motivation research demonstrates that failure often undermines goal commitment, leading people to disengage from their goals (Cochran & Tesser, 1996; Soman & Cheema, 2004). Failure has this undermining influence when people interpret it personally (Hattie & Timperley, 2007). For example, novices infer from negative feedback that they are not committed to the goal in question (Fishbach & Finkelstein, 2012), and many students interpret failure to mean they lack aptitude, which discourages subsequent goal pursuit (Yeager & Dweck, 2012). According to several motivational theories, negative feedback lowers people’s
confidence in their overall ability to pursue their goals, as well as their general expectations of success (Atkinson, 1964; Bandura & Cervone, 1983; Lewin, 1935; Weiner, 1974; Zajonc & Brickman, 1969). Only experts appear able to sustain commitment in failure’s aftermath (Louro, Pieters, & Zeelenberg, 2007).

Although past work has documented that failure undermines goal commitment and future goal pursuit, it is unclear how or whether failure affects motivation in the moment of failure itself. Past research has not explored how quickly the motivational system disengages following a failure. It is possible that failure generates an immediate motivational shutdown, undermining the individual’s motivation to attend to the task at hand. This response—“tuning out”—would imply, for example, that batters who strike out stop paying attention to the game and do not think about the way the pitch crossed the plate, the way they swung the bat, or why they struck out—which means they cannot learn from the experience.

We explored the possibility that failure—specifically, failure feedback on a task—compromises people’s motivation to learn during the failed experience itself. We predicted that the way people responded to failure feedback would undercut their ability to learn from it. Because people find failure ego threatening, they will disengage from the experience, which means they stop paying attention. Such tuning out has direct consequences for learning because people cannot learn information that they have not attended to. Because failure undermines people’s motivation to attend to a task, we hypothesized that people would learn less from failure than from success, even when failure and success were equally informative.

We tested this prediction using a novel research paradigm. In this paradigm, participants completed a learning phase, in which they guessed the correct answer to a binary-choice question (“is X ‘A’ or ‘B’?”), and then received (on the next screen) either failure feedback (“Incorrect!”) or success feedback (“Correct!”). Next, they were tested on the content of the initial questions, to see whether they learned from the feedback.

To explore why people tune out after failure feedback, we examined whether ego involvement mediated and moderated the effect of failure on learning. Specifically, we examined whether drops in self-esteem mediated the effect of negative (vs. positive) feedback on learning. We also compared how much people learned from their own failures and successes versus how much they learned from others’ failures and successes. Unlike personal failures, others’ failures do not involve the ego. Thus, we expected people to learn the same amount from others’ failures and others’ successes. It follows that people will learn more from others’ failures than from their own.

Here, we report five studies that tested whether failure undermines learning and whether ego involvement mediates and moderates this effect. First, we examined whether employees (telemarketers) learned less from failure (vs. success) when completing a learning task that was relevant to their profession. Next, we tested whether people learned less from failure (vs. success) in a learning task that contained language and relationship stimuli (i.e., learning the meaning of linguistic symbols and guessing which couples were real or not real). Across studies, we varied learning incentives in order to examine whether the inability to learn from failure generalizes across low and high levels of motivation. This range of stimuli and contexts allowed us to assess whether the failure to learn from failure is a generalizable, meaningful, real-world phenomenon.

To maximize power, we honed our measures and manipulations in pilot studies. These pilot studies yielded medium to large main effects ($d \geq 0.76$, $d = 0.65$, and $d = 0.63$; see the Supplemental Material available online). Accordingly, we targeted a minimum sample of 50 participants per cell. Power analyses conducted in G*Power software (Version 3.1; Faul, Erdfelder, Lang, & Buchner, 2007) on respective sample sizes and target alpha level (.05) revealed that power was sufficient across all studies (i.e., $\geq .80$) to detect a medium to large effect (e.g., $d = 0.60$, $\eta^2_p = .08$). Sample sizes were determined prior to data collection. All studies conducted for this research, including pilot studies, are reported in either the manuscript or the Supplemental Material (all materials and data are posted on the Open Science Framework at https://osf.io/5kbx6).

**Study 1: Not Learning From Failure in the Workplace**

In Study 1, we examined whether telemarketers at a call center learned less from failure than from success. In the learning phase, telemarketers completed multiple-choice questions about customer service—a topic of relevance to their profession. We randomly assigned them to receive success feedback on the questions they got right or failure feedback on the questions they got wrong. Both success and failure feedback contained full information on the correct answers. Nevertheless, we predicted that on a test, participants would demonstrate higher levels of learning following success than following failure.

**Method**

**Participants.** All telemarketers at a call center in the Midwest were e-mailed an invitation to complete a survey, which was described as an opportunity to gauge employee attitudes and work ethic for the purpose of
improving company culture. In this study, rather than predetermining a sample size, management invited all company employees to participate. In total, 422 telemarketers were randomly assigned to condition (success: \( n = 218 \), failure: \( n = 204 \)). A sizable portion of telemarketers (success: \( n = 46 \), failure: \( n = 47 \)) dropped out before completing the survey; however, the between-condition difference in attrition was not significant, \( \chi^2(1) = 0.23, p = .631 \). We ran analyses on the subset of telemarketers who completed the survey (success: \( n = 172 \), failure: \( n = 157 \); 63.8% female; age: \( M = 30.44 \) years, \( SD = 10.97 \)).

**Procedure.** Telemarketers read that they would be taking a survey about customers (“Today, we will ask you little-known facts about customers—what they like, dislike, how their experiences influence their attitudes to your company, etc. Your goal is to learn as much as you can”). To make the relevance of the survey salient, we reminded participants that their daily jobs are primarily about satisfying customers (“Much of your job involves pleasing customers”). Participants learned that they would answer a series of questions and that because of time constraints, they would receive performance feedback on only a few questions.

Participants then answered 10 trivia questions about customer satisfaction and customer service. Each question had two answer choices (e.g., “How much money, annually, do U.S. companies lose due to poor customer service? A. Approximately $90 billion, B. Approximately $60 billion”). The questions were based on customer-service facts taken from recent polls and research studies (Help Scout, 2018).

For this learning phase, participants were randomly assigned to a failure or a success condition. Participants in the success condition received success feedback (“Your answer was correct”) on the first four questions they answered correctly. They did not receive feedback on any other questions. If participants had fewer than four correct answers, they received feedback on just the questions answered correctly (94% received feedback on four questions, 4% received feedback on three questions, and 2% received feedback on two questions). Participants in the failure condition received failure feedback (“Your answer was incorrect”) on the first four questions they answered incorrectly. They did not receive feedback on any other questions. If participants had fewer than four incorrect answers, they received feedback on just the questions they answered incorrectly (75% received feedback on four questions, 16% received feedback on three questions, 7% received feedback on two questions, and 2% received feedback on one question). The fact that each question had just two answer choices meant that both failure and success feedback provided participants with full information on the correct answer. Success feedback communicated the correct answer directly; from failure feedback, participants could infer that the answer they did not select was the correct one. Feedback was presented on the screen that followed the participants’ choice.

We made the quiz 10 questions long and used obscure trivia facts to ensure that most participants would get 4 questions correct and 4 questions incorrect—a response distribution that would allow us to deliver similar amounts of feedback across participants in the two conditions.

Next, participants completed a distractor activity, in which they described an impactful, positive experience they had at the company. Then they entered the test phase. The test consisted of only the questions on which participants had previously received feedback (i.e., up to four questions in total). These questions had the same answer choices as the initial quiz questions, but the questions themselves were phrased in the reverse (e.g., “Which of the following amounts is NOT the amount that U.S. companies lose annually due to poor customer service? A. Approximately $90 billion, B. Approximately $60 billion”). We operationalized learning as the percentage of questions on this test that the participant answered correctly.

**Results**

Overall, participants performed slightly better than chance (50%) at identifying the correct answers to the quiz questions in the learning phase (they got 56% right, and that hit rate was similar in both conditions). As a result, participants in the failure condition received feedback and were tested on fewer questions \( (M = 3.64, SD = 0.70) \) than participants in the success condition \( (M = 3.92, SD = 0.32) \), \( t(327) = 4.75, p < .001 \). This may have made learning from failure easier than learning from success because participants in the failure condition received, had to remember, and were tested on less information.

Yet as predicted, participants in the success condition scored higher on the test \( (M = 62\% \text{ correct}, SD = 26\%) \) than participants in the failure condition \( (M = 48\% \text{ correct}, SD = 28\%) \), \( t(327) = 4.71, p < .001, d = 0.52 \), 95% confidence interval (CI) = [0.30, 0.74]. Whereas participants in the success condition learned at a rate above chance, \( t(171) = 5.84, p < .001 \), those in the failure condition did not, \( t(156) = -1.03, p = .303 \) (see Table 1 and Fig. 1).

**Study 2: Not Learning From Failure in Controlled Studies**

Although Study 1 lent preliminary support to the hypothesis that failure undermines learning motivation, it is possible that participants’ prior knowledge of the
(admittedly obscure) trivia questions biased the results. Further, it is possible that participants in the failure condition got feedback on (and thus had to learn) questions that were more difficult. Accordingly, in Study 2, we developed a script task that contained researcher-invented script symbols of which participants had no prior knowledge. We randomly assigned participants to receive success or failure feedback on the same questions, and then we tested participants to see if they learned from the feedback. Participants received a monetary bonus for each correct answer. We report four iterations of this task.

**Method**

**Participants.** Individuals of any nationality were invited to participate as long as their Amazon Mechanical Turk (MTurk) approval rating was at or above 50%. We recruited four samples. In each one, we opened the survey to 100 participants, except for the Study 2a replication, which was Table 1. Comparison of Results From the Test Phases of Studies 1, 2, 4, and 5

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Correct answers in failure condition</th>
<th>Correct answers in success condition</th>
<th>Between-conditions comparison</th>
<th>Cohen’s d</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>329</td>
<td>48% (28%)</td>
<td>62% (26%)*</td>
<td>(t(327) = 4.71, p &lt; .001)</td>
<td>0.52</td>
<td>[0.30, 0.74]</td>
</tr>
<tr>
<td>Study 2a</td>
<td>99</td>
<td>59% (41%)</td>
<td>80% (35%)*</td>
<td>(t(97) = 2.80, p = .006)</td>
<td>0.55</td>
<td>[0.15, 0.95]</td>
</tr>
<tr>
<td>Study 2a replication</td>
<td>325</td>
<td>66% (36%)*</td>
<td>88% (25%)*</td>
<td>(t(323) = 6.17, p &lt; .001)</td>
<td>0.71</td>
<td>[0.49, 0.94]</td>
</tr>
<tr>
<td>Study 2b</td>
<td>102</td>
<td>77% (31%)*</td>
<td>90% (22%)*</td>
<td>(t(100) = 2.57, p = .012)</td>
<td>0.51</td>
<td>[0.11, 0.90]</td>
</tr>
<tr>
<td>Study 2c</td>
<td>114</td>
<td>51% (44%)*</td>
<td>81% (38%)*</td>
<td>(t(112) = 3.86, p &lt; .001)</td>
<td>0.72</td>
<td>[0.34, 1.10]</td>
</tr>
<tr>
<td>Study 2d</td>
<td>103</td>
<td>67% (34%)*</td>
<td>91% (21%)*</td>
<td>(t(101) = 4.30, p &lt; .001)</td>
<td>0.85</td>
<td>[0.44, 1.25]</td>
</tr>
<tr>
<td>Study 4</td>
<td>100</td>
<td>68% (36%)*</td>
<td>88% (27%)*</td>
<td>(t(298) = 5.65, p &lt; .001)</td>
<td>0.65</td>
<td>[0.42, 0.88]</td>
</tr>
<tr>
<td>Study 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self condition</td>
<td>202</td>
<td>69% (38%)*</td>
<td>83% (33%)*</td>
<td>(F(1, 400) = 25.68, p &lt; .001)</td>
<td>0.71</td>
<td>[0.43, 1.00]</td>
</tr>
<tr>
<td>Other condition</td>
<td>200</td>
<td>80% (32%)*</td>
<td>82% (31%)*</td>
<td>(F(1, 400) = 0.81, p = .369)</td>
<td>0.04</td>
<td>[−0.04, 0.13]</td>
</tr>
</tbody>
</table>

Note: In the two columns showing the average percentage of correct answers, standard deviations are given in parentheses. Asterisks denote learning that exceeded chance. CI = confidence interval.

\(^*p \leq .001.\)
preregistered and opened to 300 participants, in accordance with a reviewer’s advice (http://aspredicted.org/blind.php?x=it2ej3). In Study 2a, MTurk returned 99 respondents (46.5% female; age: $M = 32.71$ years, $SD = 10.39$); in the Study 2a replication, MTurk returned 325 respondents (49.8% female; age: $M = 40.77$ years, $SD = 12.82$); in Study 2b, MTurk returned 102 responses (45.1% female; age: $M = 30.21$ years, $SD = 10.61$); in Study 2c, MTurk returned 114 responses (44.7% female; age: $M = 36.96$ years, $SD = 11.40$); and in Study 2d, MTurk returned 103 responses (51.5% female; age: $M = 34.98$ years, $SD = 10.16$).

**Procedure.** Prior to randomization, all participants were asked to respond to the following question in an open-text response field: “Please tell us what is your favorite book, and why.” We included this question because online participants who are not willing to invest effort tend to drop out when they see an open response question. Participants were randomly assigned to a condition only if they answered this question. Next, participants took the script task.

The script task in Study 2a consisted of three questions in Round 1 (learning phase). Each question asked participants to guess which of two symbols had a specific meaning in an invented language (e.g., “Which of the following characters in an ancient script represents an animal? $\mathcal{F}$ or $\mathcal{S}$”). Notably, as in Study 1, success and failure feedback contained equivalent amounts of information. Feedback was presented on the screen after participants chose each answer. (Note that unlike Study 1, in which participants received feedback on only some questions, participants received feedback on every question they answered.) Both types of feedback provided full information on the correct answer because each question had only two answer choices. Prior to the beginning of Round 1, participants read that performance on the later test (Round 2) was incentivized: “Whether you get the answers right or wrong in Round 1, try to learn. Round 2 will test how much you learned and bonus you $0.10 for each question you get correct!” We incentivized performance in order to ensure that participants were sufficiently motivated to learn in an online context.

After completing Round 1, participants completed a brief distractor activity—an open text box in which they reflected on their favorite music (“Tell us: what is your favorite music to listen to?”). Next, participants took a test that measured their learning. The three questions on the test (Round 2) paralleled each of the initial questions (Round 1) but were phrased in the reverse. For example, in Round 1, one question read, “Which of the following characters in an ancient script represents an animal?” On the test, participants had to answer a question with the same two symbol choices, but the question was rephrased as, “Which of the following characters represents a *non-living*, stationary object?” Because there were only two symbol choices, all participants could deduce that the symbol that was an animal in the first round was not the “non-living, stationary object” on the test, and vice versa. Thus, from the feedback, all participants received the information that would allow them to answer the test questions correctly. See the Appendix for the exact manipulation.

The Study 2a replication was a direct, preregistered replication of Study 2a. Study 2b used the same task as Study 2a, except with higher learning incentives ($\$1.50$ instead of $\$0.10$), in order to test whether participants would continue to learn less from failure under conditions that spurred higher motivation.

Study 2c used an iteration of the script task in which learning from failure required fewer mental inferences than learning from success. One reason that people may learn less from failure is that learning from failure (vs. from success) requires the participant to make more mental inferences. To learn from success, participants must remember the correct answer they were presented with, but to learn from failure, participants must infer the correct answer from the incorrect one. Accordingly, in Study 2c, we modified the test-phase questions (Round 2). In this new test phase, participants were shown the initial questions from Round 1 and were instructed to select the incorrect response to each of the questions. On this modified test, participants in the failure condition had to reselect their initial answer choices to answer correctly, whereas participants in the success condition had to infer the incorrect response on the basis of their initial correct response. Thus, failure-condition participants had to make fewer mental inferences than their success-condition counterparts to answer the test correctly.

Finally, in Study 2d, we tested whether participants would continue to learn less from failure than success when the content was social in nature. People are better at cognitive-reasoning skills when content is social (Cosmides, 1989). In Study 2d, participants completed a relationship game that was structurally similar to the script task (it featured a learning round with feedback, followed by a test), but the script symbols were replaced with social content. Each question in the relationship game asked, “Which of the following two couples are engaged?” Participants had to choose from one of two couples, following which they received feedback. Depending on condition, participants were randomly assigned to receive success feedback (“You are correct”) or failure feedback (“You are incorrect”) on each of the three questions. After the learning phase (three questions in Round 1), participants completed a short distractor activity that asked them to reflect on one of
their favorite couples. Following this, they completed a test that assessed whether they had learned which couple was engaged.

**Results**

**Study 2a.** Learning was operationalized as the percentage of Round 2 questions the participant answered correctly (out of three). The results supported our hypothesis: Participants in the failure condition ($M = 59\%$, $SD = 41\%$) learned less—that is, had fewer correct answers—than participants in the success condition ($M = 80\%$, $SD = 35\%$), $t(97) = 2.80$, $p = .006$, $d = 0.55$, 95% CI = [0.15, 0.95].

Another question is whether participants in either condition learned anything at all—that is, whether their performance exceeded chance level. Participants in the success condition learned at a level that exceeded chance (i.e., 1.5 out of 3), $t(49) = 6.06$, $p < .001$. By contrast, learning did not exceed chance in the failure condition, $t(48) = 1.45$, $p = .154$ (see Table 1 and Fig. 1).

**Study 2a replication.** As in Study 2a, participants in the failure condition ($M = 66\%$, $SD = 36\%$) learned less than participants in the success condition ($M = 88\%$, $SD = 25\%$), $t(323) = 6.17$, $p < .001$, $d = 0.71$, 95% CI = [0.49, 0.94]. Participants in the success condition learned at a level that exceeded chance, $t(165) = 19.14$, $p < .001$, as did participants in the failure condition, $t(158) = 5.82$, $p < .001$ (see Table 1 and Fig. 1).

**Study 2b.** The main effect was replicated: Participants in the failure condition ($M = 77\%$, $SD = 31\%$) learned less—they had fewer correct answers on the test—than participants in the success condition ($M = 90\%$, $SD = 22\%$), $t(100) = 2.57$, $p = .012$, $d = 0.51$, 95% CI = [0.11, 0.90]. In the success condition, participants learned at a level that exceeded chance (1.5 out of 3), $t(51) = 13.10$, $p < .001$, and the same was true in the failure condition, $t(49) = 6.07$, $p < .001$ (see Table 1 and Fig. 1). It appears that with a higher financial incentive, participants still learned from failure, albeit less than from success.

**Study 2c.** Again, the main effect was replicated: Participants in the failure condition ($M = 51\%$, $SD = 44\%$) learned less than participants in the success condition ($M = 81\%$, $SD = 38\%$), $t(112) = 3.86$, $p < .001$, $d = 0.72$, 95% CI = [0.34, 1.10]. Test scores in the success condition exceeded chance, $t(58) = 6.18$, $p < .001$; in contrast, test scores in the failure condition did not, $t(54) = 0.15$, $p = .880$ (see Table 1 and Fig. 1). Thus, in Study 2c, participants learned less from failure than from success, despite the fact that failure feedback was technically easier to learn from than success feedback because doing so required fewer mental inferences.

**Study 2d.** The main effect was replicated: Participants in the failure condition ($M = 67\%$, $SD = 34\%$) learned less than participants in the success condition ($M = 91\%$, $SD = 21\%$), $t(101) = 4.30$, $p < .001$, $d = 0.85$, 95% CI = [1.44, 1.25]. Learning in the success condition exceeded chance (2.5 out of 5), $t(50) = 13.78$, $p < .001$, as did learning in the failure condition, $t(51) = 3.50$, $p = .001$ (see Table 1 and Fig. 1). (Note that the failure condition had notably more variance than the success condition in this study, which is likely the result of ceiling effects in the success condition.)

**Study 3: Comparing Failure Feedback With No Feedback**

Studies 1 and 2 compared learning following failure with learning following success. Consequently, these studies left open the possibility that success motivates people to tune in, not that failure motivates people to tune out. In Study 3, we examined whether people learn less following failure feedback (failure condition) compared with an experience that offers no feedback at all (control condition). Thus, Study 3 tested whether participants remember fewer answers after failure feedback than after no feedback.

Measuring memory for one’s initial answer choices also allowed us to examine a different dimension of learning, which, arguably, more closely corresponds to tuning out. We expected participants who received failure feedback to have less memory of their initial answer choices than participants who received no feedback.

**Method**

**Participants.** We opened the survey to 100 participants on MTurk. Individuals of any nationality were invited to participate as long as their approval rating was at or above 50%. MTurk returned 100 responses (49% female; age: $M = 37.41$ years, $SD = 12.32$).

**Procedure.** We followed the procedure outlined in Study 2a; for example, participants were again incentivized. However, instead of comparing failure with success, we compared the failure condition with a condition that received no feedback. In the test phase (Round 2), participants saw the same multiple-choice questions from Round 1 and had to recall the answers they gave in Round 1 to each of these questions. Because we thought a test on three questions might be too easy, we expanded both Round 1 and Round 2 to include five questions.
Results

Supporting our hypothesis, results showed that participants in the failure condition (M = 59%, SD = 39%) remembered fewer of their initial answer choices than participants who received no feedback (M = 94%, SD = 16%), t(98) = 5.91, p < .001, d = 1.18, 95% CI = [0.78, 1.61]. Participants in the failure condition did not remember their answer choices at a rate above chance (2.5 out of 5), t(50) = 1.61, p = .113, whereas participants who received no feedback did remember their answer choices at a rate above chance, t(48) = 19.61, p < .001. It appears that failure—more than a no-feedback experience—led people to tune out. These results suggest that in prior studies, over and beyond any effect that success may have had on motivating people to tune in, failure led people to tune out.

Study 4: Mediation by Ego Threat

Why does failure feedback undermine learning? We hypothesized that failure hurts the ego, which leads people to tune out and not learn from the experience. In Study 4, we asked participants to report on their self-esteem following feedback. We assumed that self-reported self-esteem would capture ego threat. We hypothesized that failure feedback (vs. success feedback) would undermine self-esteem and that this would explain the lower levels of learning.

Method

Participants. We opened the experiment to 300 participants on MTurk. We recruited a slightly larger sample in Study 4, compared with Studies 2 and 3, to ensure that the study was powered to test for mediation. Individuals of any nationality were invited to participate as long as their MTurk approval rating was at or above 50%. MTurk returned 300 responses (49.3% female; age: M = 35.77 years, SD = 10.72).

Procedure. The procedure was the same as in Study 2a, with one exception: After the learning phase (Round 1), we inserted a self-report question that asked participants to report the degree to which the task undermined their self-esteem (“To what extent would you say that completing Round 1 undermined your self-esteem?”; 1 = not at all, 5 = very much). Following this question, participants completed the distractor activity and the test from Study 2a.

Results

In support of our main hypothesis, results showed that participants in the failure condition (M = 68%, SD = 36%) learned less—they had fewer correct answers—than participants in the success condition (M = 88%, SD = 27%), t(298) = 5.65, p < .001, d = 0.65, 95% CI = [0.42, 0.88]. Learning in the success condition exceeded chance (2.5 out of 5), t(152) = 17.41, p < .001, as did learning in the failure condition, t(146) = 5.96, p < .001 (see Table 1 and Fig. 1).

Next, we tested for ego threat—whether participants in the failure condition reported lower levels of self-esteem than participants in the success condition. In support of our hypothesis, results showed that participants in the failure condition felt that the task had undermined their self-esteem (M = 3.22, SD = 1.22) more than participants in the success condition did (M = 1.70, SD = 1.15), t(298) = 11.15, p < .001, d = 1.29, 95% CI = [1.04, 1.54].

Finally, we tested whether ego threat mediated the effect of failure on learning. Supporting our hypothesis, results showed that the effect of failure on learning was significantly reduced when ego threat was added to the model, R(297) = 3.24, p = .001. Ego threat mediated the indirect effect of condition on learning, β = −0.20, SE = 0.07, 95% CI = [−0.37, −0.03] in an analysis based on 10,000 bootstrap samples. In sum, participants who received failure feedback were significantly more likely than participants who received success feedback to feel that their self-esteem had been compromised. The sense that failure was ego threatening in turn undermined learning.

One limitation of Study 4 is that we relied on self-reports of ego threat. It is possible that participants did not have the insight to report their true reactions, even when we translated “ego threat” into the more familiar concept of “self-esteem.” To address this, in Study 5, we tested whether people’s ability to learn from failure improved when ego threat was removed experimentally.

Study 5: Moderation by Ego Threat—Learning From Other People’s Failures

In Study 5, we examined whether people learn from failure when ego threat is removed. We did this by comparing learning following personal successes and failures to vicarious learning following other people’s successes and failures.

Method

Participants. We opened the experiment to 400 participants. Again, individuals of any nationality were invited to participate as long as their MTurk approval rating was at or above 50%. MTurk returned 402 respondents (52.5% female; age: M = 36.21 years, SD = 11.12).
Procedure. This study used a 2 (feedback: success vs. failure; within participants) × 2 (perspective: self vs. other; between participants) mixed design. Specifically, each participant received failure feedback on one set of three questions and success feedback on another set of three questions, in counterbalanced order. Unlike in prior studies, where success and failure were between participants, participants in the current study experienced both success and failure, albeit in separate question sets.

The script task shown to the self condition was identical to the script task used in Study 2a. In the learning phase, participants answered script questions, following which they received feedback on their answer choices (Round 1). In the test phase (Round 2), we measured learning.

In contrast, in the other condition, the script task showed someone else’s performance. Prior to each set, participants in the other condition read, “In this set, you will see how someone else answered three questions and get feedback on this other person’s answers.” Participants in the other condition then clicked through Round 1 questions and saw the answer choices someone else had selected. After each answer choice, the observing participant had to reselect the answer choice, which made vicarious learning more active. Following this, the participant received feedback on the other person’s answer choice (success or failure feedback, depending on condition). He or she then completed the same test (Round 2) as participants in the self condition.

Results

A 2 (feedback) × 2 (perspective) analysis of variance revealed no main effect of perspective, $F(1, 400) = 3.17$, $p = .076$, and a main effect of feedback, $F(1, 400) = 17.75$, $p < .001$. Participants learned more from success than from failure. In support of our hypothesis, results showed a Feedback × Perspective interaction, $F(1, 400) = 8.63$, $p = .004$. Replicating the effect from prior studies, results also showed that participants learned less from personal failures ($M = 69\%$, $SD = 38\%$) than from personal successes ($M = 83\%$, $SD = 33\%$), $F(1, 400) = 25.68$, $p < .001$. However, participants learned just as much from other people’s failures ($M = 80\%$, $SD = 32\%$) as other people’s successes ($M = 82\%$, $SD = 31\%$), $F(1, 400) = 0.81$, $p = .369$.

We also calculated simple contrasts for each feedback condition. Participants learned significantly more from others’ failures than their own failures, $F(1, 400) = 9.23$, $p = .003$, but learned the same amount from personal successes and others’ successes, $F(1, 400) = 0.10$, $p = .752$. Learning was above chance level in all four cells—self-success: $t(201) = 14.45$, $p < .001$; self-failure: $t(201) = 7.28$, $p < .001$; other-success: $t(199) = 14.94$, $p < .001$; and other-failure: $t(199) = 13.39$, $p < .001$ (see Table 1 and Fig. 1). In sum, the more failure is dissociated from the self, the less people tune out, and the more they learn from failure.

We chose vicarious learning as a moderator because this moderator eliminates ego threat. That said, vicarious learning can have other effects as well—for example, it can lead people to adopt the other person’s perspective (vs. a self-perspective; Grossmann & Kross, 2014; Libby & Eibach, 2011; Pronin, Gilovich, & Ross, 2004), it can decrease overall involvement in a task (Bertsch, Pesta, Wiscott, & McDaniel, 2007), or it can prompt more abstract processing (Trope & Liberman, 2010). Nevertheless, it is difficult to see how these alternative processes could account for the results. Specifically, because perspective taking, decreasing overall task involvement, and prompting more abstract processing are less likely to differentially affect people’s ability to learn from failure (vs. success), we conclude that vicarious learning eliminated ego threat, thus increasing people’s ability to learn from failure.

General Discussion

To paraphrase the celebrated political theorist Antonio Gramsci, history teaches, but it has no pupils (Gramsci, 1977). We found that something similar happens with failure. Across five studies, participants learned less from failure feedback than from success feedback—even when both types of feedback contained full information on the correct answer. Failure feedback undermined learning motivation because it was ego threatening: It caused participants to tune out and stop processing information.

Our findings advance motivation theory and, in particular, past theoretical and empirical work that argues that negative feedback undermines goal commitment (Atkinson, 1964; Bandura & Cervone, 1983; Cochran & Tesser, 1996; Fishbach & Finkelstein, 2012; Hattie & Timperley, 2007; Lewin, 1935; Soman & Cheema, 2004; Weiner, 1974; Yeager & Dweck, 2012; Zajonc & Brickman, 1969). Complementing this past work, which describes how failure affects motivation in the future, our studies explored how failure feedback affects motivation in the present—the moment of failure itself. Our key result is that people find failure feedback ego threatening, which leads them to tune out and stop processing information. In other words, failure undermines learning. It is possible that these immediate effects underlie the longer-term demotivating effects of failure on goal commitment. Tuning out from a pursuit in the moment of failure could be the first step in a chain reaction that distances and discourages people from the goal they are pursuing.
It is possible that this tune-out reaction depends on the size of the failure. In the well-documented phenomenon of aversion learning, animals that taste poison, receive shocks, or experience other “large” failures learn to avoid these threats in the future (Garcia, Lasiter, Bermudez, & Deems, 1985). It is possible that for large failures, the attentional pull of the negative experience overrides the motivation to tune out. Nevertheless, there are many large failures—for example, failures in close relationships—that people might be bad at detecting over long periods of time, despite their size and importance. Similarly, many failures are small (e.g., a failed experiment that a researcher discards as noninformative), yet they accumulate a significant amount of information that people might fail to learn from.

We found that people struggle to learn from failure feedback in the field, using a task that presented employees with relevant professional information, and in online samples, using tasks involving language and relationship stimuli. We found the effect in both the United States and in the United Kingdom, though these cultures are admittedly similar. It is still an open question whether people’s failure to learn from their mistakes would generalize to individuals in other cultures—and in particular, to individuals in cultures that have different attitudes to failure. For example, Japanese individuals persist longer after they fail than after they succeed, whereas Americans do the opposite (Heine et al., 2001). Thus, it is unclear whether Japanese individuals, like the American and British participants in our samples, would learn less from failure than from success. Another open question is whether certain failures are more easily taught than others. Future research is needed to determine whether the observed effect generalizes when feedback is more personalized, more detailed, or delivered in a different way (i.e., by a caring mentor). We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

Our results have practical implications. People who want to learn may be better able to do so via successful experiences than via unsuccessful experiences. When failure feedback is inevitable, our results suggest that people will learn more if failure feedback can be separated from the ego. No matter the precise method for reducing ego involvement—for example, positioning people as vicarious learners or instructing people to reappraise feedback in less threatening terms—our results suggest that reducing the degree to which failure involves the ego will promote learning.

Appendix

This appendix contains the text for the manipulations used in Study 2a and the Study 2a replication. The manipulations for the other studies are posted at https://osf.io/5bx6.

<table>
<thead>
<tr>
<th>Failure condition</th>
<th>Success condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welcome! Today you will answer some language questions about a researcher-manufactured ancient script. When you click to the next page, you will begin Round 1. Whether you get the answers right or wrong in Round 1, try to learn. Round 2 will test how much you learned and bonus you $0.10 for each question you get correct! QUESTION #1. Which of the following characters in an ancient script represents an animal? <img src="image" alt="Character Options" /></td>
<td>YOU ANSWERED THIS QUESTION INCORRECT! YOU ANSWERED THIS QUESTION CORRECT!</td>
</tr>
<tr>
<td>QUESTION #2. Which of the following characters in an ancient script represents a person. <img src="image" alt="Character Options" /></td>
<td>YOU ANSWERED THIS QUESTION INCORRECT! YOU ANSWERED THIS QUESTION CORRECT!</td>
</tr>
</tbody>
</table>
Action Editor
Michael Inzlicht served as action editor for this article.

Author Contributions
Both authors developed the study concept and study designs. L. Eskreis-Winkler collected and analyzed the data. L. Eskreis-Winkler drafted the manuscript; A. Fishbach provided critical revisions. Both authors approved the final version of the manuscript for submission.

ORCID iD
Lauren Eskreis-Winkler https://orcid.org/0000-0003-1977-5677

Acknowledgments
We are grateful to the participants and to Carman Fowler, who made this work possible. We thank Shane Frederick and lab members for indispensable feedback. L. Eskreis-Winkler extends special thanks to Ari Lustig, who learns from all his successes—he has no failures.

Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material
Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619881133

Open Practices
All data and materials have been made publicly available via the Open Science Framework and can be accessed at https://

### Example Table

<table>
<thead>
<tr>
<th>Failure condition</th>
<th>Success condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUESTION #3. Which of the following characters in an ancient script represents a bird?</td>
<td></td>
</tr>
<tr>
<td>☻ ☾ ☿</td>
<td></td>
</tr>
</tbody>
</table>

YOU ANSWERED THIS QUESTION INCORRECT! YOU ANSWERED THIS QUESTION CORRECT!

Thanks for answering those initial questions. Right now take a short breather. Tell us: what is your favorite music to listen to?

This is Round 2.

Based on what you learned about the researcher-manufactured ancient script in Round 1, answer the final three questions below. For each question you get correct, you will earn bonus cash ($0.10).

QUESTION #1. Which of the following characters represents a non-living, stationary object?

○ ☻ ☾ ☿

QUESTION #2. Which of the following characters represents a non-living, stationary object?

○ ☾ ☿ ☻

QUESTION #3. Which of the following characters represents a non-living, stationary object?

○ ☿ ☾ ☻
osf.io/5kbx6. The design and analysis plans for the Study 2a replication were preregistered at http://aspredicted.org/blind.php?x=it2ej3. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619881133. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Notes

1. We began the survey with this question in all online panel studies reported in this article (Studies 2–5), and we randomly assigned participants to a condition only if they provided an answer.

2. Nevertheless, results were consistent when feedback was presented on the same page, alongside the answer choices. See Study S4 in the Supplemental Material.

References


