Not Learning from Failure—the Greatest Failure of All

Lauren Eskreis-Winkler
Ayelet Fishbach
University of Chicago

In press, Psychological Science

Author Note
To whom correspondence should be addressed: Lauren Eskreis-Winkler, Booth School of Business, University of Chicago, 5807 S. Woodlawn Avenue, Chicago, IL 60637. Phone: 610-329-7514. Email: leskreiswinkler@uchicago.edu. Or Ayelet Fishbach, Booth School of Business, University of Chicago, 5807 S. Woodlawn Avenue, Chicago, IL 60637. Fax: 773-753-0556. Phone: 773-834-8673. Email: ayelet.fishbach@ChicagoBooth.edu.
Abstract
Our society celebrates failure as a teachable moment. Yet in five experiments \((N = 1,674)\), 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, since there were only two answer choices. However, on a follow-up test, participants learned less from failure feedback than success feedback. This effect replicated across professional, linguistic and social domains—even when learning from failure was less cognitively taxing and when learning was incentivized. Those 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 personal success, yet they learned just as much from others’ failure as others’ success. Thus, when the ego concerns are muted, people tune in and learn from failure.

Keywords: learning; feedback; failure; ego-threat; motivation
Not Learning from Failure — The Greatest Failure of All

Our society celebrates failure as a teachable moment (Brown, 2014; Maxwell, 2007; Stoker & Hindle, 2003). Research appears to support the point. People react more strongly—physiologically, cognitively, and emotionally—to negative events than positive ones (Baumeister et al., 2001; Kahneman & Tversky, 1979; Rozin & Royzman, 2001; Taylor, 1991), in ways that arguably enhance learning. For example, compared to positive stimuli, negative stimuli command more attention (Ohman, 2007; Ohman, Flykt, & Esteves, 2001; 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.

We explore the alternative: that people learn less from failure. Whether people learn from failure depends not only on whether failure is attention-grabbing, but on peoples’ motivation to attend to it. If people are motivated to ignore their failures, then they will not attend to them, and will not learn. 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 peoples’ 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).

While 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. Possibly, failure generates an immediate motivational shutdown, undermining the individual’s motivation to attend to the task at hand. This tuning out response would imply, for
example, that a batter who strikes out at the plate stops paying attention to the game, does not think about the way the pitch crossed the plate, the way she swung, or why she struck out—which means she cannot learn from the experience.

We explore the possibility that failure—specifically, failure feedback on a task—compromises peoples’ motivation to learn during the failed experience itself. We predict that the way people respond to failure feedback undercuts their ability to learn from it. Because people find failure ego-threatening, they will disengage from the experience, which means they stop paying attention, or, tune out. Tuning out has direct consequences for learning, since people cannot learn information that they have not attended to. Insofar as failure undermines peoples’ motivation to attend to a task, we hypothesize that people will learn less from failure than success, even when failure and success are 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 received (on the next screen) failure feedback (“Incorrect!”) or success feedback (“Correct!”). Next, they were tested on the content in the initial questions, to see if they learned from the feedback.

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

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) learn less from failure (vs. success) when completing a learning task that is relevant to their profession. Next, we tested whether people learn less from failure (vs. success) in a learning task that contains language and relationship stimuli (i.e., learning the meaning of linguistic symbols and guessing which couples are real vs. not). Across studies, we varied learning incentives in order to examine whether the inability to learn from failure (vs. success)
Not Learning from Failure

generalizes across low and high levels of motivation. This range of stimuli and contexts allows us to assess whether the failure to learn from failure is a generalizable, meaningful, real-world phenomenon.

To maximize power, across studies we honed our measures and manipulations in pilot studies. These pilot studies yielded medium-to-large main effect sizes \( d = .76, d = .65, \) and \( d = .63; \) see SOM-R). Accordingly, we targeted a minimum sample of 50 participants per cell. Power analyses conducted in G*Power on respective sample sizes and target alpha level \( (\alpha = .05) \) revealed that power was sufficient across all studies (i.e., \( \geq .80 \)) to detect a medium to large effect (e.g., \( d = .60, \) \( \eta^2_p = .08 \)). Sample sizes were determined prior to data collection. All studies conducted for this research, including pilot studies, are collectively reported in the manuscript and SOM-R (full materials and data are posted on OSF: https://osf.io/5kbx6).

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

In Study 1, we examined whether telemarketers at a call center services company learned less from failure than success. In the learning phase, telemarketers completed multiple-choice questions about customer service—a topic of relevance to their profession. We randomized 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 versus failure.

**Method**

**Participants.** All telemarketers at a call center services company in the midwest were emailed an invitation to complete a survey, 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, \( N = 422 \) telemarketers were randomized to condition \( (n_{success} = 218; n_{failure} = 204) \). A sizable portion of telemarketers \( (n_{failure} = 47, n_{success} = 46) \) dropped out before completing the survey; however, there was not differential attrition, \( \chi^2(1) = 0.23, p = .631 \). We ran analyses on the subset of telemarketers who completed the survey \( (n_{success} = 172; n_{failure} = 157; 63.8\% \) female; \( M_{age} = 30.44, SD_{age} = 10.97 \)).

**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, participants were reminded 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 due to time constraints, they would only receive performance feedback on a few questions.

Participants then answered ten trivia questions about customer satisfaction and customer service. Each question had two answer choices (e.g., “How much money, annually, do US 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 (see https://www.helpscout.net/75-customer-service-facts-quotes-statistics/).

For this learning phase, participants were randomized 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 (on all other questions they received no feedback). They did not receive feedback on any other questions. If a participant had fewer than four correct answers, they received feedback on just the questions they answered correctly (94% received feedback on four questions, 4% received feedback on three questions, 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 (on all other questions they received no feedback). 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 on two questions, and 2% received feedback on one question). Of key importance, each question had just two answer choices. This means 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 four questions correct and four 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 the telemarketers had previously received feedback (so, 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 US companies lose annually due to poor customer service? A. Approximately $90 billion, B. Approximately $60 billion”). We operationalized learning as the percent of questions on this test that the participant answered correctly.

**Results**

Overall, participants were slightly better than chance at identifying the correct answers to the quiz questions in the learning phase (they got 56% right, and that hit rate was similar across 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 (vs. success) easier, because participants in the failure condition received, had to remember, and were tested on, less information.

![Learning by Condition (Studies 1-2 & Studies 4-5).](image)

*Figure 1*. Learning by Condition (Studies 1-2 & Studies 4-5). Across versions of the task that tested learning [Studies 1-2 & Study 4, Study 5 (Self)], participants learned less from failure than success. However, when the failures and successes were someone else’s, participants learned no less from failure than success (Study 5, Other). Error bars represent one standard error.
Yet to the contrary, as predicted, telemarketers in the success condition scored higher on the test ($M = 62\%$ correct, $SD = 26\%$) than telemarketers in the failure condition ($M = 48\%$ correct, $SD = 28\%$), $t(327) = 4.71, p < .001, d = .52, 95\%$ CI [.30, .74]. Whereas the success condition learned at a rate above chance ($50\%$), $t(171) = 5.84, p < .001$, the failure condition did not, $t(156) = -1.03, p = .303$. See Table 1 and Figure 1 for a summary of results in Study 1, and all other experiments that measured learning.
Table 1.

Results Summary for Studies that Measured Learning (Studies 1-2 & 4-5).

Note. Asterisks denote learning that exceeded chance. *** p ≤ .001.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Average (SD) % of Correct Answers Post Failure</th>
<th>Average (SD) % of Correct Answers Post Success</th>
<th>Did participants in the failure condition learn less?</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>.52</td>
<td>.30, .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>.55</td>
<td>.15, .95</td>
</tr>
<tr>
<td>Study 2a—Rep</td>
<td>325</td>
<td>66% (36%)***</td>
<td>88% (25%)***</td>
<td>t(323) = 6.17, p &lt; .001</td>
<td>.71</td>
<td>.49, .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>.51</td>
<td>.11, .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>.72</td>
<td>.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>.85</td>
<td>.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>.65</td>
<td>.42, .88</td>
</tr>
<tr>
<td>Study 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self</td>
<td>202</td>
<td>69% (38%)***</td>
<td>83% (33%)***</td>
<td>F(1, 400) = 25.68, p &lt; .001</td>
<td>.71</td>
<td>.43, 1.00</td>
</tr>
<tr>
<td>Other</td>
<td>200</td>
<td>80% (32%)***</td>
<td>82% (31%)***</td>
<td>F(1, 400) = 0.81, p = .369</td>
<td>.04</td>
<td>-.04, .13</td>
</tr>
</tbody>
</table>

Study 2: Not Learning From Failure in Controlled Experiments

While Study 1 lends 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 was 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 which contained researcher-invented script symbols of which participants had no prior knowledge. We randomized participants to receive success or failure feedback on the same questions, and then 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. Participants of any nationality were invited to participate, so long as their MTurk approval rating was at or above 50%. Sequentially, we recruited four samples. In each one, we opened the survey to 100 participants (except Study 2a-Replication, which was preregistered and opened to 300 participants; http://aspredicted.org/blind.php?x=it2ej3, per
reviewer’s advice). In Study 2a, MTurk returned 99 respondents (46.5% female; $M_{age} = 32.71$, $SD_{age} = 10.39$); in Study 2a-Replication MTurk returned 325 respondents (49.8% female; $M_{age} = 40.77$, $SD_{age} = 12.82$); in Study 2b, MTurk returned 114 responses (44.7% female; $M_{age} = 36.96$, $SD_{age} = 11.40$); in Study 2c MTurk returned 102 responses (45.1% female; $M_{age} = 30.21$, $SD_{age} = 10.61$); and in Study 2d MTurk returned 103 responses (51.5% female; $M_{age} = 34.98$, $SD_{age} = 10.16$).

**Procedure.** Prior to randomization, all participants were asked an open-text response question: “…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. We began the survey with this question in all online-panel studies reported in the paper (Studies 2-5), and only randomized participants who provided an answer. 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? ⊳ or ⊲”). Notably, as in Study 1, success and failure feedback contained equivalent amounts of information. Feedback was presented on the screen after participants made each answer choice.¹ (Note that unlike Study 1, in which participants received feedback on only some questions, in this and all subsequent studies, 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 this study, and all subsequent studies in the manuscript, 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

¹ Nevertheless, results are consistent when feedback is presented on the same page, alongside the answer choices. See Study 4S in SOM-R.
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 re-phrased to: “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.

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 people may learn less from failure is because learning from failure (vs. success) requires the participant to make more mental inferences. To learn from success, the participant has to remember the correct answer they were presented with, but to learn from failure, the participant has to infer the correct answer from the incorrect one. Accordingly, in Study 2c, we modified the test phase questions (Round 2). This new test phase showed participants the initial questions from Round 1, and instructed them to select the incorrect response to each of the questions. On this modified test, participants in the failure condition had to re-select their initial answer choices to answer correctly, whereas participants in the success condition had to infer the incorrect response based on their initial correct response. Thus, failure 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” which 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 randomized to receive success (“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 which asked them to reflect on one of their favorite couples. Following this, they completed a test which assessed whether they had learned which of the couples were engaged.

Results.

Study 2a. Learning was operationalized as the percent of Round 2 questions the participant answered correct (out of three). The results supported our hypothesis: participants in the failure condition ($M = 59\%, SD = 41\%$) learned less—they had fewer correct answers—than participants in the success condition ($M = 80\%, SD = 35\%$), $t(97) = 2.80, p = .006, d = .55, 95\% CI = [.15, .95]$.

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 (a chance score is 50\%, i.e., 1.5 out of 3), $t(49) = 6.06, p < .001$. By contrast, learning did not exceed chance in the failure condition, $t(48) = 1.45, p = .154$.

Study 2a-Replication. As in Study 2a, participants in the failure condition ($M = 66\%, SD = 36\%$) learned less than participants in the success condition ($M = 88\%, SD = 25\%$), $t(323) = 6.17, p < .001, d = .71, 95\% CI = [.49, .94]$. Participants in the success condition learned at a level that exceeded chance (a chance score is 50\%, i.e., 1.5 out of 3), $t(165) = 19.14, p < .001$, as did participants in the failure condition, $t(158) = 5.82, p < .001$.

Study 2b. The main effect 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 ($M = 90\%, SD = 22\%$), $t(100) = 2.57, p = .012, d = .51, 95\% CI [.11, .90]$. The success condition learned at a level that exceeded chance (a chance score is 50\%, i.e., 1.5 out of 3), $t(51) = 13.10, p < .001$, as did the failure condition, $t(49) = 6.07, p < .001$. It appears that with a higher financial incentive, participants still learned from failure, albeit less than from success.

Study 2c. The main effect replicated: participants in the failure condition ($M = 51\%, SD = 44\%$) learned less—they had fewer correct test answers—than participants in the success condition ($M = 81\%, SD = 38\%$), $t(112) = 3.86, p < .001 d = .72, 95\% CI = [.34, 1.10]$. Test
scores in the success condition exceeded chance (50%), \( t(58) = 6.18, p < .001 \); in contrast, test scores in the failure condition did not, \( t(54) = 0.15, p = .880 \). Thus, in Study 2c, participants learned less from failure than success—despite the fact that failure feedback was technically easier to learn from than success feedback, since doing so required fewer mental inferences.

**Study 2d.** The main effect replicated: participants in the failure condition (\( M = 67\%, SD = 34\% \)) learned less—they had fewer correct answers—than participants in the success condition (\( M = 91\%, SD = 21\% \)), \( t(101) = 4.30, p < .001, d = 0.85, 95\% CI = [.44, 1.25] \). Learning in the success condition exceeded chance (50% or 2.5/5), \( t(50) = 13.78, p < .001 \), as did learning in the failure condition, \( t(51) = 3.50, p = .001 \). (Note that the failure condition has 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 to No Feedback**

Studies 1-2 compared learning following failure to learning following success. As such, these studies leave 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 to an experience that offers no feedback at all (control condition). Thus, Study 3 tests whether participants remember fewer answers after failure feedback than a no-feedback experience.

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 (vs. no feedback) to have worse memory for their initial answer choices.

**Method**

**Participants.** We opened the survey to 100 participants on MTurk. Participants of any nationality were invited to participate, so long as their approval rating was at or above 50%. MTurk returned 100 responses (49.0% female; \( M_{age} = 37.41, SD_{age} = 12.32 \)).

**Procedure.** We followed the procedure outlined in Study 2a (for example, as in Study 2a, participants were incentivized). However, instead of comparing failure to success, the failure condition was compared to 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 3 questions might be too easy, we expanded both Round 1 and Round 2 to include 5 questions.

**Results**

Supporting our hypothesis, 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 = [.78, 1.61]. Participants in the failure condition did not remember their answer choices at a rate above chance (50% or 2.5/5), $t(50) = 1.61, p = .113$, 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 (vs. success) feedback would undermine self-esteem, and that this would explain lower levels of learning.

**Method**

**Participants.** We opened the experiment to 300 participants on MTurk. We recruited a slightly larger sample in Study 4, as compared to Studies 2-3, in order to ensure we were powered to test for mediation. Participants of any nationality were invited to participate, so long as their MTurk approval rating was at or above 50%. MTurk returned 300 responses (49.3% female; $M_{age} = 35.77, SD_{age} = 10.72$).

**Procedure.** The procedure was the same as Study 2a; the one exception was that after the learning phase (Round 1), we inserted a self-report question which 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, 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 (50\% or 2.5/5), $t(152) = 17.41, \ p < .001$, as did learning in the failure condition, $t(146) = 5.96, \ p < .001$.

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, participants in the failure condition felt the task had undermined their self-esteem ($M = 3.22, \ SD = 1.22$) more than participants in the success condition ($M = 1.70, \ SD = 1.15$), $t(298) = 11.15, \ p < .001, \ d = 1.29, \ 95\% \ CI = [1.04, \ 1.54]$.

Finally, we tested whether ego threat mediated the effect of failure on learning. Supporting our hypothesis, the effect of failure on learning was significantly reduced when ego-threat was added to the model, $t(297) = 3.24, \ p = .001$. Ego-threat mediated the effect of condition on learning, $\beta_{\text{indirect}} = -0.20, \ SE = .07, \ 95\% \ CI [-0.37, \ -0.03]$; based on 10,000 bootstrap samples. In sum, participants who received failure (vs. success) feedback were significantly more likely 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. Possibly, participants did not have the self-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 peoples’ ability to learn from failure improved when ego-threat was removed experimentally.

**Study 5: Moderation by Ego Threat—Learning from Other Peoples’ Failures**

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

**Method**

**Participants.** We opened the experiment to 400 participants. Participants of any nationality were invited to participate, so long as their MTurk approval rating was at or above 50\%. MTurk returned 402 respondents (52.5\% female; $M_{\text{age}} = 36.21, \ SD_{\text{age}} = 11.12$).
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-subjects, in the current study, participants 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 re-select the answer choice him/herself, 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). They then completed the same test (Round 2) as participants in the “self” condition.

Results

An ANOVA of Feedback × Perspective 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 failure. In support of our hypothesis, we found a Feedback × Perspective interaction, $F(1, 400) = 8.63, p = .004$. Replicating the effect from prior studies, participants learned less from personal failures ($M = 69\%, SD = 38\%$) than personal successes ($M = 83\%, SD = 33\%$), $F(1, 400) = 25.68, p < .001$. However, participants learned just as much from others’ failures ($M = 80\%, SD = 32\%$) as others’ 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 success and others’ success, $F(1, 400) = 0.10, p = .752$. 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, $t(199) = 14.94, p < .001$; other/failure, $t(199) = 13.39, p < .001$. In sum, the more failure is removed from the self, the less people tune out, and the more they learn from it.

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’s perspective (vs. a self-perspective; Grossmann & Kross, 2014; Libby & Eibach, 2011; Pronin, Gilovich, & Ross, 2004), can decrease overall involvement in a task (Bertsch, Pesta, Wiscott, & McDaniel, 2007), or prompt more abstract processing (Trope & Liberman, 2010). Nevertheless, it is difficult to see how these alternative processes could account for the results. Specifically, since perspective taking, decreasing overall task involvement, or prompting more abstract processing, are less likely to differentially affect peoples’ ability to learn from failure (vs. success), we conclude that vicarious learning eliminated ego threat, thus increasing peoples’ ability to learn from failure.

**General Discussion**

The celebrated political theorist, Antonio Gramsci, said, “history teaches, but it has no pupils” (Gramsci, 1977). We find something similar happens with failure. Across five studies, participants learned less from failure than 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 which 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, we explore 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 miss the information it offers. In other words, failure undermines learning. It is possible that these immediate effects underlie the longer-term de-motivating 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 who 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. It is also the case that many failures are small (e.g., a failed experiment that a researcher discards as non-informative), 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, in 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 US and in the UK, though these cultures are admittedly similar. It is still an open question whether peoples’ failure to learn from failure 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 individuals in our samples, would learn less from failure than success. Another open question is whether certain failures are more teachable 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, versus unsuccessful, experiences. When failure feedback is inevitable, our results suggest people will learn more if failure feedback can be removed from the ego. No matter the precise method for reducing ego-involvement—for example, positioning people as vicarious learners, or instructing people to re-appraise 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 text for the manipulation for Study 2a and Study 2a-Replication. The exact manipulations for the other studies are posted on: https://osf.io/5kbx6.

<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.</td>
<td></td>
</tr>
<tr>
<td>When you click to the next page, you will begin Round 1.</td>
<td></td>
</tr>
<tr>
<td>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!</td>
<td></td>
</tr>
</tbody>
</table>

QUESTION #1. Which of the following characters in an ancient script represents an animal?
- □ ṭ
- □ ṁ

YOU ANSWERED THIS QUESTION INCORRECT!

YOU ANSWERED THIS QUESTION CORRECT!

QUESTION #2. Which of the following characters in an ancient script represents a person.
- □ ū
- □ ū

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).
Author Contributions

Both authors developed the study concept and study designs. L. Eskreis-Winkler performed data collection and analysis. L. Eskreis-Winkler drafted the manuscript; A. Fishbach provided critical revisions. Both authors approved the final version of the manuscript for submission.
References


