AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2024, Vol. 153, No. 9, 2328–2344
https://doi.org/10.1037/xge0001632

Learning From Failure: The Roles of Self-Focused Feedback, Task Expectations, and Subsequent Instruction

Sebahat Gok[1, 2] and Emily R. Fyfe[3]

[1] Cognitive Science Program, The College of Arts and Sciences, Indiana University-Bloomington
[2] Department of Instructional Systems Technology, Indiana University-Bloomington
[3] Department of Psychological and Brain Sciences, Indiana University-Bloomington

Previous research indicates that failure feedback leads people to *tune out* from the task, which is detrimental to their learning (Eskreis-Winkler & Fishbach, 2019; Keith et al., 2022). The current work aims to identify ways to optimize learning from failure feedback. We conducted six preregistered experiments ($N = 1{,}306$) to replicate and extend the findings from Eskreis-Winkler and Fishbach (2019) with novel tests of self-focused feedback, task expectations, and subsequent instruction. The detriments of failure feedback were replicated in Studies 1a, 1b, and 1c, which altered the focus of the feedback message to be self-focused (e.g., your answer) or task-focused (e.g., the answer). The detriments of failure feedback were also replicated in Study 2 when the task expectations were manipulated to easy versus hard. These results generally underscored the robustness of the results from the original study. However, Study 3 established boundary conditions. When it was a rule-based task and brief instructions on the rule were provided after feedback, there was no evidence of a detrimental effect of failure, and failure feedback, in some conditions, resulted in even better learning than success feedback for learning new material. We conclude that the *tune-out* reactions to failure during feedback disappear and may even be reversed when subsequent learning opportunities are provided.

Public Significance Statement

This study investigates the influence of failure feedback—messages indicating incorrect responses—on learning. Our experiments reveal that individuals tend to learn less from failure feedback compared to success feedback on memorization tasks. This trend holds across variations in feedback wording (“you are incorrect” vs. “that is incorrect”) and regardless of the expected difficulty of the task, suggesting that failure feedback can demotivate learners at the moment it is received. However, our experiments also find that the detrimental effects of failure feedback vanish when subsequent learning opportunities are provided. Moreover, under some conditions, initial failure feedback promotes greater learning on a subsequent task relative to initial success feedback. Given the ubiquity of failure feedback in real-life scenarios, it may be important to design learning experiences that enable learners to monitor their performance and provide opportunities for continued engagement.

Keywords: feedback valence, task expectations, failure, learning, productive failure

Feedback is defined as the information an agent provides regarding one’s performance or understanding of a task, and it is one of the most powerful instructional factors that improve learning (Hattie & Timperley, 2007; Mandouit & Hattie, 2023). The learning gains caused by feedback have often been attributed to its informational value. For example, it is theorized that feedback improves learning by providing the information that signals the gap between the actual and desired performance (Hattie & Timperley, 2007) and by providing the opportunity to retrieve relevant information from long-term memory (Butler & Woodward, 2018). Yet, the amount of relevant information in the feedback message does not always reliably relate to learning gains (e.g., Fyfe & Rittle-Johnson, 2016), which suggests there are factors that influence feedback’s effectiveness beyond how much information it provides. To investigate some

Laura E. Thomas served as action editor.
Sebahat Gok https://orcid.org/0000-0001-5057-4668
The preregistration of all experiments reported in this article is available online on the Open Science Framework at https://osf.io/j85ep/registrations. The materials and de-identified data for all experiments are also available on the Open Science Framework at https://osf.io/j85ep/ (Gok & Fyfe, 2024). The portions of this work were previously disseminated by the authors at the 45th Annual Conference of the Cognitive Science Society, the 2023 American Educational Research Association Meeting, and the 44th Annual Conference of the Cognitive Science Society.

Sebahat Gok played a lead role in data curation, formal analysis, investigation, and writing–original draft and an equal role in methodology and project administration. Emily R. Fyfe played a lead role in funding acquisition and writing–review and editing and an equal role in methodology, project administration, and supervision.
Correspondence concerning this article should be addressed to Sebahat Gok, Cognitive Science Program, The College of Arts and Sciences, 1001 East 10th Street Indiana University-Bloomington, IN, 47405. Email: sebahgok@gmail.com

of these factors, the current series of experiments focused on its valence—feedback on failures versus successes—and features of the learning environment that may improve learning from failure. In particular, we examined the self-focused nature of the feedback message, the task expectations, and subsequent learning opportunities.

The Role of Failure Feedback

For over a century, feedback researchers have attempted to answer the question of how to improve the effectiveness of feedback. Earlier behaviorist studies framed feedback as a reinforcement stimulus that increased or decreased the likelihood of a response (Skinner, 1938, as cited in Butler & Woodward, 2018). Failure feedback (i.e., feedback that indicates an answer is incorrect) was viewed as a punishment while success feedback (i.e., feedback that indicates an answer is correct) was viewed as a reward, and punishment was considered less effective than reinforcement. Accordingly, behaviorists avoided giving failure feedback and instead designed instructional environments with heavily prompted questions to guide learners to the correct answer and ultimately to success feedback (for a review, see Butler & Woodward, 2018). By the 1970s, however, many behaviorist ideas had not lived up to empirical evidence. Empirical studies showed feedback was more beneficial when it followed incorrect rather than correct responses (Anderson et al., 1971; Gilman, 1969; Guthrie, 1971). Accordingly, feedback was conceptualized as information that facilitated the correction of errors (Kulhavy, 1977). Based on this empirical evidence, and based on the fact that errors are ubiquitous in learning, this view has largely dominated the field and has enforced a strong emphasis on failure feedback.

However, a recent series of randomized experiments by Eskreis-Winkler and Fishbach (2019)[1] has challenged the notion that feedback is more effective after errors. Across eight independent samples and experiments, they found that adults learned dramatically less from failure feedback than from success feedback. This result was true even when the prior knowledge of the task was controlled and even when learning from failure feedback required fewer mental inferences than success feedback. The authors suggested that the mechanism was motivational; the ego threat invoked by failure feedback caused people to "tune out and stop processing information," which was detrimental to their learning (Eskreis-Winkler & Fishbach, 2019, p. 8).

This explanation is consistent with a range of theoretical ideas and empirical evidence. For example, researchers have long noted the inherent dilemma of failure feedback—noting that "though in many situations it may be very useful to obtain feedback for its error corrective and uncertainty reducing properties, individuals may be reluctant to actively pursue it in an attempt to protect their self-esteem" (Ashford & Cummings, 1983, p. 377). More recently, Grundmann et al. (2021) formalized these ideas in their *Model of Motivated Feedback Disengagement*, arguing that failure feedback, in particular, may produce negative emotions (e.g., shame, disappointment) that cause learners to actively avoid the task in some way. These ideas also echo earlier ethnographic studies from educational settings, which point out learners may dismiss failure feedback to protect their self-worth (e.g., Eva et al., 2012; Sargeant et al., 2009).

Despite these complexities, failure feedback is pervasive in the real world, and it is critically important to identify ways to learn from it optimally. Research in this area can have important theoretical implications for better specifying how people process negative information and also practical implications for designing feedback messages in effective ways. A group of researchers recently replicated the EW-F study and identified two boundary conditions to the tune-out effect caused by failure feedback (Keith et al., 2022). Accordingly, the effect diminishes when the incentives for performance are framed in terms of a loss (e.g., losing money for each wrong answer) rather than a gain (e.g., earning money for each correct answer) and when the feedback message contains the correct answer in addition to a correct/incorrect judgment.

In the current work, we replicate and extend the EW-F study by investigating three other potential boundary conditions: the self-focused nature of the feedback message (Study 1), the task expectations (Study 2, Study 3), and subsequent learning opportunities (Study 3). These features were primarily motivated by a recommendation in EW-F that suggested "people will learn more if failure feedback can be separated from the ego" (Eskreis-Winkler & Fishbach, 2019, p. 9). All experiments were preregistered on the Open Science Framework at https://osf.io/j85ep/registrations, and we replicated sample size choices by EW-F.

Feedback Focus: Self- Versus Task-Focused Feedback

A long-standing theory on feedback interventions suggests that feedback has the power to direct the learner's attention in different ways (Kluger & DeNisi, 1998). This theory suggests that feedback that directs attention to the self will be less effective than feedback that directs attention to the task. Attention to the self can activate affective reactions that consume cognitive resources and diminish the intended effect of feedback. Accordingly, if one perceives a threat to the self, the gap pointed out by the feedback may be resolved by allocating less attention to the task or abandoning it altogether to attain a positive self-view. In support of this notion, an early meta-analysis found that feedback interventions that contained cues to the self (e.g., praise, verbal feedback involving the salience of another person) yielded lower performance than feedback messages that did not (Kluger & DeNisi, 1996). These results have led to recommendations to deliver feedback in nonevaluative, neutral ways (e.g., Schartel, 2012).

One factor that may influence the evaluative nature of the feedback and whether the learner can "separate the ego from the feedback" is the use of personal pronouns in the feedback message (e.g., you). Self-focused personalization may enhance the saliency of the learner and potentially increase the use of internal attributions (e.g., "I got that wrong because I am not smart."). Prior research has examined this self-focused language in the context of students learning from text passages. The second-person pronouns (e.g., you) relate the learner to the material more directly by using the self as a reference point, whereas the corresponding third-person pronouns (e.g., he) treat the learner as a more detached observer (Son & Goldstone, 2009). In some contexts, these modest manipulations of personalized pronouns (e.g., "Imagine you are a doctor") in text passages were found to hinder learning relative to neutral ones (e.g., "Imagine a doctor"; Son & Goldstone, 2009) but benefitted learning in others (e.g., "your lung" vs. "the lung"; Moreno & Mayer, 2004).

[1] For conciseness, we refer to Eskreis-Winkler and Fishbach's (2019) reference as EW-F in the rest of this article.

However, to our knowledge, these manipulations have not been examined in the context of the feedback message.

In the study by EW-F, the feedback messages were self-focused: “You answered this question correct/incorrect.” We postulated that the participants did not learn from failure feedback because the use of the second-person pronoun (“You”) might have focused their attention on the self in a negative, ego-threatening way. We hypothesized that more task-focused language (“The answer was correct/incorrect”) would help separate failure feedback from the ego and make the learning difference between success and failure feedback smaller.

Task Expectations

Another factor that may influence whether learners can “separate the ego from the feedback” is the learner’s expectations about the difficulty of the task—whether they think the task will be easy or hard. Expecting the task to be hard—and thus expecting to fail—may result in a self-fulfilling prophecy by which the learner believes the task is unachievable, puts in minimal effort, receives failure feedback, and uses that negative, evaluative information to confirm their original beliefs (Mangels et al., 2012). In this case, the expectation of failure leads the learner to interpret the negative feedback in an ego-threatening way (e.g., “I knew I wouldn’t be able to do this task”) and produces low performance. Certainly, there have been some arguments in favor of setting expectations for failure (e.g., DePasque & Tricomi, 2015; Norem & Cantor, 1986; Sargeant et al., 2009); the idea is that it can be a protective mechanism in risky situations and minimize the intensity of negative emotions in the case of failure. However, experimental evidence tends to suggest that expecting the task to be hard—and therefore setting expectations for failure—is particularly detrimental for learners receiving failure feedback (e.g., Fyfe & Brown, 2020), and expecting the task to be easy may help those learners succeed.

In the study by EW-F, task expectations were fairly neutral, but leaned in the direction of expecting the task to be hard. For example, in Study 1, the directions said: “Today, we will ask you little known facts about customers … your goal is to learn as much as you can.” Indicating that the facts are “little known” suggests participants are likely to get them wrong. We postulated that setting expectations for failure might have focused attention on the self in an ego-threatening way. We explored whether setting expectations for success would help separate failure feedback from ego and make the learning difference between success and failure feedback smaller.

Failure’s Influence on Subsequent Instruction

A third factor that may influence whether learners can “separate the ego from the feedback” is the availability of further learning opportunities. Whereas the findings from EW-F suggest that failure feedback leads to tune-out reactions *initially*, other lines of research have shown delayed benefits of failure on tasks that are removed from the immediate ego-threatening experience of failure (e.g., Kapur, 2008; Schwartz et al., 2011; Schwartz & Martin, 2004). For example, in Kapur (2008), one group of students was assigned to solve difficult, ill-structured physics problems, and the other group was assigned easier, well-structured ones. Initially, the first group generated poorer performance with lower quality solutions (i.e., they experienced failure). However, the trend was reversed on a subsequent task: Students who had initially worked on ill-structured problems and failed outperformed those who had started with simpler tasks and succeeded. These findings suggest latent benefits of failure, which are revealed when the learner is provided with subsequent learning opportunities and then assessed. This phenomenon, sometimes referred to as “productive failure” (Kapur & Bielaczyc, 2012), is theorized to be particularly beneficial in preparing learners for subsequent instruction by activating their prior knowledge and aiding them in discerning the critical features of the task.

Though productive failure is more generally about the benefits of engaging with difficult tasks, there are reasons to hypothesize that failure feedback, in particular, can have motivational benefits for tuning into a subsequent task. For example, the *Model of Motivated Feedback Disengagement* (Grundmann et al., 2021) suggests that while people may initially disengage from failure feedback to fulfill hedonic goals (e.g., to feel good at the moment), those who want to improve their performance will, over time, regulate their emotions and adopt an engagement strategy to learn from the task. Conversely, not having such initial failure experiences might create an illusion of competence, leading to a lower level of arousal and disengagement from subsequent tasks (e.g., Wagner et al., 2024).

In the study by EW-F, learning was assessed only for the task on which participants received the feedback. We postulated that the participants did not learn from failure feedback because they were not provided any subsequent learning opportunities that were removed from the immediate ego-threatening experience. We explored whether failure feedback would have delayed benefits by preparing participants to learn from subsequent instruction.

In summary, our goal was to replicate and extend the findings by EW-F. We explored whether three factors—feedback focus, task expectations, and subsequent instruction—would moderate the negative influences of failure feedback on learning. To foreshadow the results, we did not find evidence that feedback focus or task expectations independently moderated these effects. However, we found that subsequent instruction did so under some conditions.

Transparency and Openness

All experiments were preregistered. All de-identified raw data, analytic methods (code), experimental scripts, design and analysis plan, study preregistration, and analysis plan registrations are available on the Open Science Framework at https://osf.io/j85ep/?view_only=8a89dfd7b6104cceb9deae4f8a390b70.

Study 1: Self-Focused Versus Task-Focused Feedback

In Study 1, we replicated EW-F Study 2a and extended it with a new variable: feedback focus. Our extension consisted of modest wording manipulations of the feedback message in an attempt to shift the focus from the self to the task across three independent experiments.

Study 1a: Method

In the EW-F Study 2a, the participants received self-focused feedback that indicated either success or failure (“You answered this question correct!/incorrect!”). In our Study 1a, we included these two conditions to replicate the EW-F results. We also added two new conditions with task-focused feedback (“The answer is correct!/incorrect!”). Thus, Study 1a consisted of four conditions with two independent variables: feedback valence (success or failure) and feedback focus (self-focused or task-focused). Participants were randomly assigned to one of the four conditions. An immediate posttest measured participants' learning. We expected the effect of feedback valence would be smaller for conditions with task-focused feedback.

Participants and Design

We determined the sample size prior to data collection based on the original EW-F report. In the EW-F Study 2a, they included 99 participants across two conditions. A power analysis using G*Power3 (Faul et al., 2007) suggests a sample size of 100 yields a power of $1 - \beta = .80$ to detect a medium effect size of $d = .60$ and $\eta^2 = .08$, with a Type 1 error probability of $\alpha = .05$. Because we had four between-subject conditions, we aimed to double the sample size and recruit approximately 200 people. We had 206 participants initiate the study, and three of them failed an attention check that occurred prior to the assignment of any conditions. The final participants were 203 undergraduate students from a large university in the Midwestern United States who received credit in their psychology course. Their average age was 19.0 years ($SD = 1.8$). Most students reported their gender as female (74%), followed by male (22%), nonbinary (four people), transgender (two people), and two chose not to report. Most reported their ethnicity as White (65%), followed by Asian/Asian American (15%), Black/African American (10%), Hispanic/Latino (8%), and mixed (2%). Finally, about half were freshman (53%), followed by sophomore (25%), junior (13%), and senior (9%).

The study had a 2 (feedback valence: success or failure) × 2 (feedback focus: task or self) between-subjects design with random assignment producing similar sample sizes across the four conditions (self-focused success feedback: $n = 53$, each remaining group: $n = 50$).

Procedure

The materials and procedure were identical to those in the EW-F Study 2a. The original materials in EW-F can be found at https://osf.io/5kbx6.

The study took place as a single online session on Qualtrics. It started with an open-ended attention-check question that aimed to eliminate participants who were unwilling to invest effort. Then, Round 1 (the learning phase) began, which consisted of three multiple-choice questions. The questions had participants guess which of two symbols had a particular meaning in a researcher-invented script (e.g., "Which of the following characters in an ancient script represents an animal?"). Each question was followed by a new screen that displayed a stand-alone feedback message. Given the arbitrary nature of the symbols, there were no objectively correct answers and so the valence of the feedback could be manipulated. Some participants received self-focused success feedback on each answer ("You answered this question correct!"), and others received self-focused failure feedback ("You answered this question incorrect!"). These two conditions are identical to those in the EW-F study. In two additional conditions, some participants received task-focused success feedback on each answer ("The answer was correct!"), and others received task-focused failure feedback ("The answer was incorrect!"). Participants received the same feedback after each question given that this was a between-subject manipulation meaning they received all success feedback or all failure feedback.

After Round 1, the participants responded to a brief distraction task and then completed Round 2 (the test phase), which also consisted of three multiple-choice questions. The test phase questions were identical to the learning phase questions except that they were worded in reverse with superordinate categories. For example, if the learning question asked which symbol represents an animal, then the test question asked which symbol represents a stationary, nonliving object and presented the same two symbol choices. Thus, Round 2 tested whether participants learned the correct responses to the questions in Round 1.

The present study replicated the procedures of the previous EW-F Study 2a with the following variations. First, the present study recruited participants from an undergraduate subject pool while the EW-F study recruited participants from Amazon Mechanical Turk (Mturk). Second, the present study rewarded all participants with course credit regardless of their performance while the EW-F study rewarded bonus payments for each correct answer in Round 2. Third, the present study included two additional conditions to manipulate feedback focus.

Study 1a: Results

We operationalized learning as the percentage of the Round 2 test questions the participants answered correctly as in the EW-F study. See Table 1 and Figure 1 for the descriptive statistics as a function of condition. For our analyses, we first replicated the analyses in the EW-F study. The conditions of self-focused success feedback and self-focused failure feedback in the present study correspond to the success and failure conditions in the EW-F study.

We replicated the finding that self-focused failure feedback led to significantly lower learning ($M = 69\%$, $SD = 29\%$) than self-focused success feedback ($M = 93\%$, $SD = 19\%$), $t(83) = 4.86$, $p < .001$, 95% CI [.14, .33], $d = .97$.[2]

Second, we tested the effects of the novel feedback focus variable by conducting a 2 (feedback focus: task or self) × 2 (feedback valence: success or failure) analysis of variance (ANOVA) predicting percentage correct on the Round 2 test questions. There was a significant main effect of feedback valence as success feedback ($M = 94\%$, $SD = 18\%$) resulted in better learning than failure feedback ($M = 72\%$, $SD = 29\%$), $F(1, 199) = 40.98$, $p < .001$, $\eta^2 = .17$. There was not a significant main effect of feedback focus, $F(1, 199) = 1.45$, $p = .22$, $\eta^2 = .00$, and in contrast to our hypothesis, there was no significant interaction effect between feedback valence and feedback focus, $F(1, 199) = .30$, $p = .58$, $\eta^2 = .00$.

Study 1b: Method

In Study 1a, we replicated the main finding from EW-F that people learn less from failure feedback than from success feedback. Contrary to our hypothesis, changing the focus of the feedback from self to task did not significantly reduce the discrepancy between success and failure conditions. In Study 1b, we replicated the four

[2] We also compared the learning performance of each group to chance level (50%) based on our preregistered analysis plan. All groups performed significantly better than chance level.

Table 1
Comparison of Results Across Studies 1, 2, and 3

Condition	*N*	Correct answers in success condition *M* (*SD*)	Correct answers in failure condition *M* (*SD*)	Between-conditions comparison	Cohen's *d*	95% CI
			Study 1a			
Self-focused	103	93% (19%)	69% (29%)	$t(83) = 4.86, p<.001$	.97	[.14, .33]
Task-focused	100	95% (18%)	75% (29%)	$t(81) = 4.13, p<.001$	.83	[.10, .30]
			Study 1b			
Self-focused	161	76% (38%)	69% (35%)	$t(158) = 1.22, p = .22$	.19	[−.04, .18]
Task-focused	163	72% (40%)	62% (39%)	$t(161) = 1.58, p = .11$	.25	[−.02, .22]
			Study 1c			
Self-focused	55		45% (34%)	$t(106) = -1.03, p = .30$	−.17	[−.20, .06]
Task-focused	54		52% (37%)			
			Study 2			
Expect-easy	104	86% (28%)	63% (35%)	$t(96) = 3.75, p<.001$	.73	[.11, .35]
Expect-hard	103	71% (35%)	64% (37%)	$t(101) = .99, p = .32$	.20	[−.07, .21]
			Study 3a			
Rule 1 (identical items)						
Expect-easy	113	92% (21%)	88% (27%)	$t(82) = .95, p = .35$	.19	[−.04, .13]
Expect-hard	105	88% (23%)	97% (11%)	$t(65) = -2.58, p = .01$	−.53	[−.17, −.02]
Rule 1 (isomorphic items)						
Expect-easy	113	90% (22%)	91% (24%)	$t(96) = -.08, p = .93$	−.02	[−.09, .08]
Expect-hard	105	90% (21%)	92% (21%)	$t(101) = -.38, p = .70$	−.07	[−.09, .06]
Rule 2 (novel task)						
Expect-easy	113	66% (28%)	76% (28%)	$t(100) = -1.95, p = .05$	−.37	[−.21, .00]
Expect-hard	105	69% (30%)	80% (27%)	$t(95) = -1.91, p = .05$	−.38	[−.21, .00]
			Study 3b			
Rule 1 (identical items)						
Expect-easy	135	90% (21%)	91% (18%)	$t(128) = -.39, p = .69$	−.07	[−.08, .05]
Expect-hard	110	81% (31%)	85% (26%)	$t(101) = -.67, p = .50$	−.13	[−.14, .07]
Rule 1 (isomorphic items)						
Expect-easy	135	90% (22%)	91% (24%)	$t(111) = -.12, p = .90$	−.02	[−.08, .07]
Expect-hard	110	78% (29%)	85% (25%)	$t(102) = -1.32, p = .19$	−.26	[−.17, .03]
Rule 2 (novel task)						
Expect-easy	135	71% (30%)	69% (29%)	$t(119) = .37, p = .71$	.07	[−.08, .12]
Expect-hard	110	55% (34%)	71% (28%)	$t(100) = -2.64, p<.01$	−.50	[−.28, −.04]

Note. CI = confidence interval.

conditions from Study 1a with two independent variables: feedback valence (success or failure) and feedback focus (self-focused or task-focused). But in Study 1b, we extended the context of the study by recruiting participants using Amazon Mechanical Turk. We also introduced variations to incentivize learning from errors with bonus payments and by providing the question and responses along with the feedback messages (see Figure 2). We also included an ego-threat measure from EW-F Study 4 to test whether higher ego threat mediated the relation between condition and learning.

Participants and Design

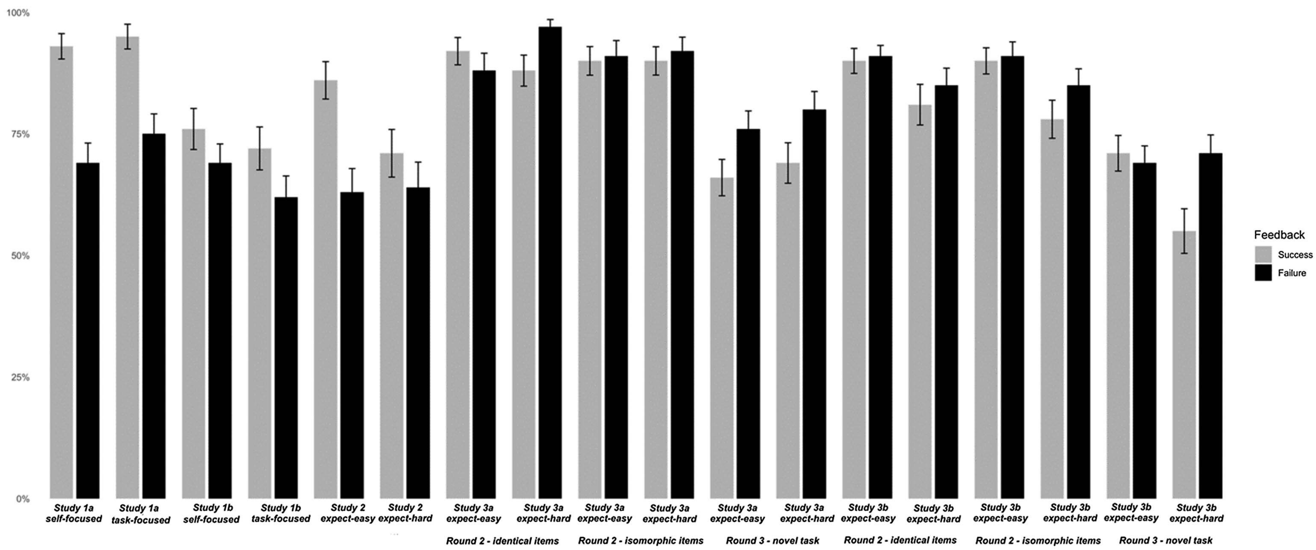
We based our sample size choice on EW-F's Study 4 ($n = 300$) to ensure our study was fully powered to test for mediation. We had 351 participants initiate the study, and 21 of them failed an attention check that occurred prior to the assignment of the condition. Six additional participants were excluded from the analyses because they did not answer all questions (task-focused failure: $n = 2$, task-focused success: $n = 2$, and self-focused failure: $n = 2$). The final participants were 324 adults recruited on Amazon Mechanical Turk (MTurk) from the United States whose approval rating was at or above 50%. Most reported their gender as male (66%), followed by female (33%), with one person choosing not to report. Most reported their ethnicity as White (75%), followed by Black/African American (8%), Asian/Asian American (7%), Native American (5%), Hispanic/Latino (3%), Native Hawaiian/Pacific Islander (two people), and mixed (one person). Their highest educational degree obtained was undergraduate (61%), graduate (21%), high school (18%), and middle school (one person). Their reported average age was 38 ($SD = 10$).

The study had a 2 (feedback valence: success or failure) × 2 (feedback focus: task or self) between-subjects design with random assignment producing similar sample sizes across the four conditions (self-focused success: $n = 81$, self-focused failure: $n = 80$, task-focused success: $n = 82$, task-focused failure: $n = 81$).

Procedure

The materials and procedures were identical to Study 1a with the following alterations. First, the feedback was not displayed in isolation; instead, the question, options, and the participant's

Figure 1
Comparison of Results Across Studies 1, 2, and 3



Note. Average percentage of correct answers in the test phases of Studies 1 and 2 as a function of the success and failure feedback conditions. Errors represent ±1 *SE*. *SE* = standard error.

original response also appeared on the screen for their review (see Figure 1). Second, after completing the Round 1 learning phase, the participants were presented with a 5-point Likert scale question: "To what extent would you say that completing Round 1 undermined your self-esteem?" (1 = *not at all*, 5 = *very much*). As in EW-F Study 4, the participant's response to this question was treated as their level of ego threat. Third, each correct answer from the Round 2 test phase was rewarded with a \$0.10 bonus payment as was done in EW-F Study 2a.

Study 1b: Results

First, we replicated the analyses in the EW-F study. We compared learning between self-focused success and self-focused failure groups. In this study, self-focused failure feedback led to descriptively lower learning (M = 69%, SD = 35%) than self-focused success feedback (M = 76%, SD = 38%), but this difference was not statistically significant, $t(158) = 1.22$, $p = .22$, 95% CI [−.04, .18], $d = .19$.[3]

Second, we tested the novel effects of feedback focus. The results were similar to that of Study 1a. The 2 (feedback focus: task or self) × 2 (feedback valence: success or failure) ANOVA predicting percentage correct on the Round 2 test questions revealed a significant main effect of feedback valence as success feedback (M = 73%, SD = 38%) resulted in better learning than failure feedback (M = 65%, SD = 37%); $F(1, 320) = 3.96$, $p = .004$, $\eta^2 = .01$. There was not a significant main effect of feedback focus, $F(1, 320) = 2.03$, $p = .15$, $\eta^2 = .00$, and there was no significant interaction, $F(1, 320) = 0.11$, $p = .73$, $\eta^2 = .00$.

Finally, we tested whether there was a difference between the ego-threat levels of the groups. We failed to replicate the EW-F result that failure feedback produced higher levels of ego threat. A 2 × 2 ANOVA predicting ego threat showed that no factor had a significant effect: $F_{\text{valence}}(1, 319) = 1.30$, $p = .25$, $\eta^2 = .00$; $F_{\text{focus}}(1, 319) = .07$, $p = .79$, $\eta^2 = .00$; $F_{\text{valence::focus}}(1, 319) = .58$, $p = .44$, $\eta^2 = .00$. Across all four groups, self-reported ego threat was quite high: task-focused failure—$M = 3.36$, $SD = 1.22$; self-focused failure—$M = 3.18$, $SD = 1.48$; task-focused success—$M = 3.03$, $SD = 1.71$; and self-focused success—$M = 3.12$, $SD = 1.66$. Further, higher ego threat was significantly negatively correlated with learning, $r(298) = -.28$, $p < .001$.

Study 1c: Method

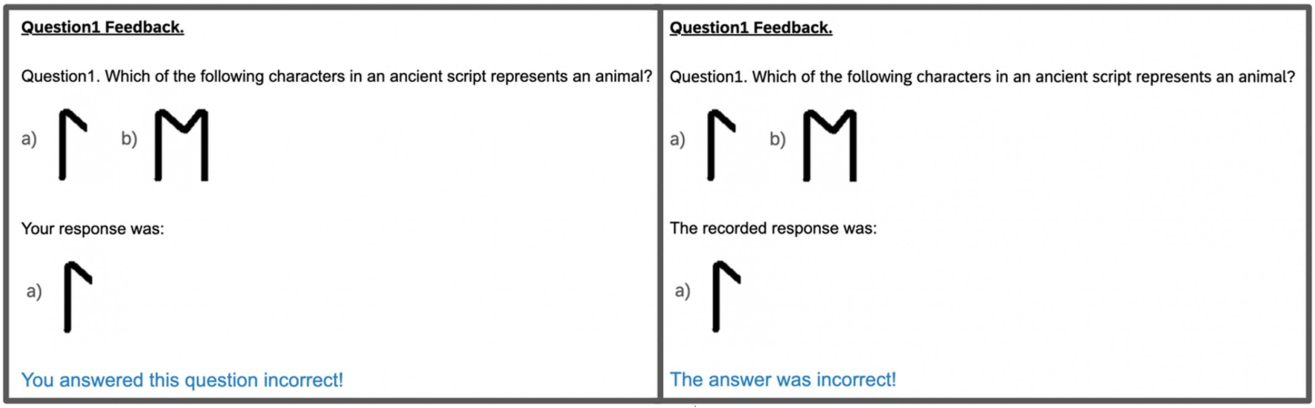
The results from Study 1b largely mirrored those of Study 1a; when considering all four conditions, there was evidence that people learned less from failure feedback than from success feedback, and changing the focus of the feedback (e.g., your answer vs. the answer) did not influence learning. In contrast to EW-F, we found no evidence that ego threat explains the detrimental effects of failure feedback. In Study 1c, we investigated the role of feedback focus one final time by adopting a more extreme wording for self-focused feedback ("You were incorrect!" as opposed to "You answered this question incorrect!"). We only included failure feedback conditions and compared this self-focused feedback to task-focused feedback.

Participants and Design

We determined the sample size based on the effect size from our prior pilot study. A power analysis using G*Power3 (Version 3.1; Faul et al., 2007) yielded the result of sample size 110 with power of $1 - \beta = .80$, the effect size of $d = .54$ and $\eta^2 = .08$, and Type 1 error probability of $\alpha = .05$. We had 119 participants initiate the

[3] We also compared the learning performance of each group to chance level (50%) based on our preregistered analysis plan. All groups performed significantly better than chance level.

Figure 2
Study 1b Example Feedback Message



Note. The figure shows a sample question from the Round 1 learning session for self-focused failure (on the left) and task-focused failure (on the right). The success conditions had an identical setup with only the wording change from "incorrect" to "correct." See the online article for the color version of this figure.

study, and six of them failed an attention check that occurred prior to the assignment of condition. Four additional participants were excluded from analyses because they did not answer all questions (task-focused failure: $n = 3$, and self-focused failure: $n = 1$). Final participants were 109 adults recruited on MTurk from the United States whose approval rating was at or above 50%. Most reported their gender as male (58%), followed by female (42%). Most reported their ethnicity as White (79%), followed by Asian/Asian American (8%), Black/African American (8%), Hispanic/Latino (two people), Native American (two people), and mixed (one person). Their highest educational degree obtained was undergraduate (75%), high school (17%), and graduate (8%). Their reported average age was 36 ($SD = 10$). The participants were randomly assigned to receive either self-focused failure feedback ($n = 55$) or task-focused failure feedback ($n = 54$).

Procedure

The materials and procedures were the same as in Study 1b with a few exceptions. The self-focused feedback was more direct (i.e., "You were incorrect!"), and the feedback message was once again provided as a single stand-alone message as in Study 1a. We decided to keep the feedback slide simple to revert to a direct replication of EW-F. Also, we provided failure feedback to all participants, rather than assigning some to receive success feedback.

Study 1c: Results

We tested the novel effect of feedback focus, and once again, there were no condition differences. A between-subjects t test showed that there was not a significant difference in learning scores between self-focused failure feedback ($M = 45\%$, $SD = 34\%$) and task-focused failure feedback ($M = 52\%$, $SD = 37\%$), $t(106) = -1.03$, $p = .30$, CI [−.20, .06], $d = -.17$. In fact, neither group performed significantly different than chance level, self-focused failure: $t(54) = 1.12$, $p = .26$, 95% CI [.35, .54], $d = -.15$; task-focused failure: $t(53) = 0.36$, $p = .71$, 95% CI [.41, .61], $d = .08$. Also, a between-subjects t test showed that there was not a significant difference in ego-threat levels between the self-focused failure feedback group ($M = 3.49$, $SD = 1.12$) and the task-focused failure feedback group ($M = 3.55$, $SD = 1.26$), $t(105) = -.28$, $p = .77$, CI [−.52, .39], $d = -.17$. There was a nonsignificant negative correlation between higher ego-threat levels and learning, $r(107) = -.08$, $p = .38$.

Study 1: Discussion

Study 1 aimed to replicate an experiment by Eskreis-Winkler and Fishbach (2019) with the extension of a new variable, the focus of feedback. Participants were randomly assigned to receive success feedback or failure feedback, with self-focused language (e.g., you) or task-focused language (e.g., the answer). Overall, we replicated the main finding that people learn less from failure feedback than from success feedback with textual fact-based information (i.e., the meaning of a new symbol). These findings challenge the long-held assumption that feedback is most helpful in correcting errors (Butler & Woodward, 2018; Kulhavy, 1977).

We failed to replicate one key finding from the EW-F study. The EW-F study found that the group receiving failure feedback reported significantly higher levels of ego threat compared to the success feedback group. In our Study 1b, however, we did not observe significant differences in ego threat across groups, even though ego threat negatively correlated with learning. However, we do not conclude that ego threat does not mediate the relationship between feedback valence and learning. The results in our Study 1b may be due to the minor change we implemented. For example, displaying the options, participants' responses, and feedback messages on the same page might have reduced the salience of feedback valence. In our Study 1b, the effect size of feedback valence is smaller relative to EW-F's experiment. In general, the average learning performance in the success groups was lower in our study (around 73%), compared to the same subject pool in EW-F's experiments, where it ranged from 80% to 91%. These minor deviations may have made learning more difficult overall, making it hard to detect differences in self-reported ego-threat levels across the conditions.

Regarding feedback focus, we did not find evidence in support of our hypothesis that task-focused feedback would facilitate learning relative to self-focused feedback, especially in the failure feedback condition. Given that we tested this manipulation in three independent studies, we conclude that feedback focus does not appear to influence learning from feedback in this specific context (i.e., novel text-based memory task on a computer). However, it is certainly possible that the use of personal pronouns matters in other contexts. For example, prior research suggests that feedback provided by a computer is less threatening than feedback provided by a human supervisor (Kluger & Adler, 1993), and using in-person self-focused feedback may matter more. Also, previous studies have detected the influence of similar word manipulations in computer-based learning settings, such as Moreno and Mayer (2004) and Son and Goldstone (2009), but, in the present study, attention to self may have mattered less in a novel text-based learning situation in which no prior knowledge is expected or possible.

Study 2: Easy Versus Hard Task Expectations

In Study 2, we continued our replication efforts of the EW-F study and extended it with a new variable: task expectations. Our extension consisted of modifying the task instructions to explicitly indicate whether the task should be easy or hard. Prior to the Study 2 reported here, we conducted several pilot studies to test manipulations that effectively altered participants' expectations. Among those trials, the most pronounced differences in expectations occurred when instructions were related to mathematics and accompanied by pictures of equations (see Figure 3). Therefore, in Study 2, we reformulated the task to be equations with researcher-invented scripts (as opposed to assigning meaning to symbols). The questions included equations with novel arbitrary symbols, and as in previous experiments, there was no objectively correct answer. We experimentally manipulated success and failure feedback, and the task simply required memorizing the correct option as in the EW-F paradigm.

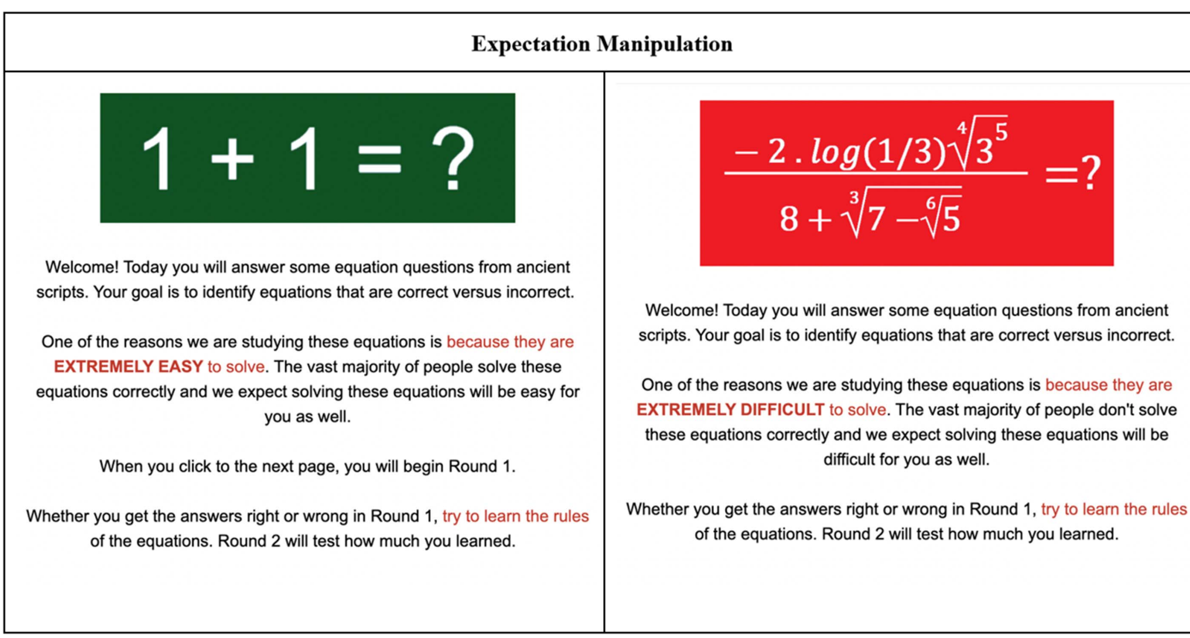
Study 2: Method

In Study 2, we included success feedback and failure feedback conditions to replicate the EW-F results. However, we changed the task instructions to explicitly manipulate whether participants expected the task to be easy or hard. Thus, Study 2 consisted of four conditions with two independent variables: feedback valence (success or failure) and task expectations (easy or hard). We expected the effect of feedback valence would be smaller for conditions with easy task expectations—in which participants expect to succeed.

Participants and Design

Because we had four between-subject conditions, we modeled our sample size on Study 1a and aimed for a sample of approximately 200 adults. We had 230 participants initiate the study, and 14 of them failed an attention check that occurred prior to the assignment of condition. Five additional participants were excluded from analyses because they did not answer all questions (easy and failure: $n = 4$, hard and success: $n = 1$). Final participants

Figure 3
Task Expectation Manipulation



Note. Participants were randomly assigned to expect-easy (on the left) or expect-hard (on the right) conditions. The script was adapted from "This is easy, you can do it! Feedback during mathematics problem solving is more beneficial when students expect to succeed," by E. R. Fyfe and S. A. Brown, 2020, *Instructional Science*, *48*(1), pp. 32–33 (https://doi.org/10.1007/s11251-019-09501-5). See the online article for the color version of this figure.

included 207 adults recruited on MTurk from the United States whose approval rating was at or above 50%. Most reported their gender as male (70%) followed by female (30%). Most reported their ethnicity as White (75%) followed by Hispanic/Latino (7%), Native American (7%), Asian/Asian American (6%), Black/African American (3%), mixed race (two people), and Native Hawaiian/Other Pacific Islander (one person). Their highest educational degree obtained was bachelor's (72%), followed by high school (14%), graduate (13%), and middle school (two people). Their reported average age was 36 (SD = 10).

The study had a 2 (feedback valence: success or failure) × 2 (task expectations: expect-easy or expect-hard) between-subjects design with random assignment producing similar sample sizes across the four conditions (easy-success: n = 53, hard-success: n = 52, easy-failure: n = 51, hard-failure: n = 51).

Procedure

The procedures were identical to Study 1b with the following alterations. Before Round 1 started, we manipulated participants' expectations about the task (see Figure 3). All participants read that they are going to study equations from an ancient script and identify equations that are correct or incorrect. In the expect-easy condition, they were told that this is extremely easy and most people learn them correctly. In the expect-hard condition, they were told that this is extremely difficult and most people do not learn them correctly. Then, a manipulation check was administered in which participants were asked to rate how easy or difficult they expected the task to be on a 5-point Likert scale. This was followed by a Round 1 learning phase and a Round 2 test phase that were similar to the previous experiments, except the questions were now about equations with symbols instead of the meaning of symbols (see Figure 4). For the three questions during the learning phase, participants had to select the symbol that correctly satisfied the equation. Because the equations were arbitrary, we manipulated whether participants received failure feedback or success feedback on all three questions. For the three questions during the test phase, we reversed the wording to be consistent with the prior experiments (i.e., "Which of the following equations in an ancient script is INCORRECT?").

Study 2: Results

First, we tested whether our expectation manipulation worked by running a 2 × 2 ANOVA predicting their expectation ratings out of five, with higher scores representing more difficulty. The manipulation was successful. There was a significant main effect of expectation condition, $F(1, 203) = 58.69$, $p < .01$, $\eta^2 = .22$, as participants in the hard conditions (hard-success: M = 3.46, SD = 1.19; hard-failure: M = 3.35, SD = 1.38) expected the task to be significantly more difficult than participants in the easy conditions (easy-success: M = 2.38, SD = 1.09; easy-failure: M = 1.98, SD = .88). Given that the expectation manipulation occurred before any feedback was provided, there was no main effect of feedback valence, $F(1, 203) = 2.48$, $p = .12$, $\eta^2 = .01$, or valence by expectation interaction, $F(1, 203) = .80$, $p = .37$, $\eta^2 = .00$.

Second, we tested the novel effects of task expectations on the learning scores.[4] We conducted a 2 (feedback valence: success or failure) × 2 (task expectations: easy or hard) ANOVA predicting percentage correct on the Round 2 test questions. Similar to the prior experiments, there was a main effect of feedback valence, with those receiving success feedback (M = 78%, SD = 33%) outperforming those who received failure feedback (M = 63%, SD = 36%); $F(1, 203) = 10.38$, $p < .001$, $\eta^2 = .05$. There was not a main effect of task expectations, $F(1, 203) = 2.09$, $p = .15$, $\eta^2 = .01$, and the interaction of valence and expectation was not significant, $F(1, 203) = 2.97$, $p = .08$, $\eta^2 = .01$.

Third, we tested whether there was a difference between the ego-threat levels of the groups. A 2 × 2 ANOVA with ego-threat ratings as the dependent variable revealed no main effect of feedback valence, $F(1, 203) = 0.06$, $p = .79$, $\eta^2 = .00$; no main effect of task expectations, $F(1, 203) = 0.00$, $p = .99$, $\eta^2 = .00$; and no interaction, $F(1, 203) = 1.74$, $p = .10$, $\eta^2 = .01$. All four groups reported fairly high levels of ego threat on the scale from 1 to 5 (easy-success: M = 3.20, SD = 1.52; hard-success: M = 3.46, SD = 1.47; easy-failure: M = 3.41, SD = 1.18; hard-failure: M = 3.13, SD = 1.33).

Study 2: Discussion

The findings of Study 2 indicate that we successfully manipulated learners' expectations, with some expecting the task to be easy and others expecting the task to be hard. However, these varying expectations did not influence learning, and contrary to our hypothesis, the negative impact of failure feedback persisted even after manipulating task difficulty expectations. These findings generally replicate the detriments of failure feedback reported in EW-F. However, there were some descriptive trends suggesting that task expectations may matter, so we continue to manipulate task expectations in Study 3. Additionally, in Study 3, we altered the study design to resemble more authentic learning tasks. Specifically, we introduced a rule-based task with objectively correct answers and provided subsequent learning opportunities.

Study 3: Failure Feedback's Influence on Future Tasks

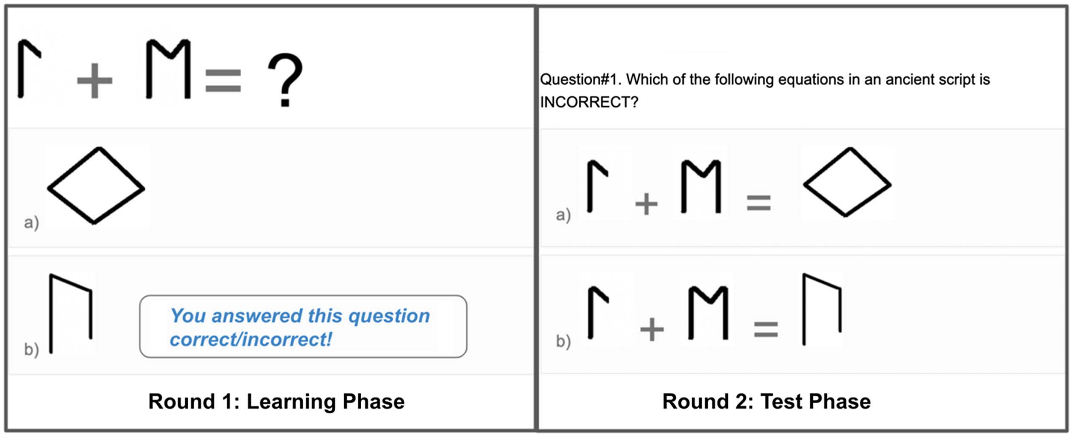
Study 3a: Method

Across the experiments so far, learning has been based on text-based memorization—essentially linking arbitrary symbols to a meaning or to an equation. These experiments have consistently replicated the main detrimental influence of failure feedback, demonstrating the robustness of the original results. Study 3 extends the context to a rule-based task. In this type of task, feedback on one trial might influence the learner's performance on the next trial as the feedback provides relevant information about the rule governing the whole set of items.

In addition, Studies 1 and 2 only assessed learning of the three specific items on which learners received feedback. However, in more authentic contexts, there are often subsequent items or tasks on which to apply the knowledge one learns. In Study 3, the participants receive brief instruction after the feedback and are tested on the items on which they received feedback (Round 2), but also on novel items governed by the same rules (Round 2) as well as a set of novel items governed by a new set of rules that they are briefly taught (Round 3). Thus, Study 3 systematically dissociates the test content from the initial task on which feedback is given.

[4] We also compared each single group to chance level. All groups performed significantly better than chance level.

Figure 4
Example Questions in Study 2



Note. See the online article for the color version of this figure.

Participants and Design

Because we had two factors and four cells, we modeled our sample size on Study 1a and aimed for a sample of approximately 200 participants. We had 219 participants initiate the study, and one of them failed an attention check that occurred prior to the assignment of the condition. The final participants were 218 undergraduate students from a large university in the Midwestern United States who received credit in their psychology course. Their average age was 19.05 years (SD = 1.80). Most students reported their gender as female (55%), followed by male (44%) and transgender (one participant). Most reported their ethnicity as White (67%), followed by Asian/Asian American (12%), Black/African American (11%), Hispanic or Latino (7%), and Other (3%). Their year in college was freshman (49%), followed by sophomore (34%), junior (10%), and senior (6%).

Overall, we attempted to mimic the between-subjects study design from Study 2a with two factors: task expectations (easy or hard) and feedback valence (success or failure). Initially, participants were randomly assigned to one of the two expectation conditions (expect-hard: n = 113, and expect-easy: n = 105). Because we used a rule-based task with objectively correct answers, we had to modify how we assigned participants to receive success or failure feedback. The participants solved the three questions during Round 1, and based on their actual performance, they were post hoc categorized as low performers who received failure feedback (n = 104) or high performers who received success feedback (n = 114).

Procedure

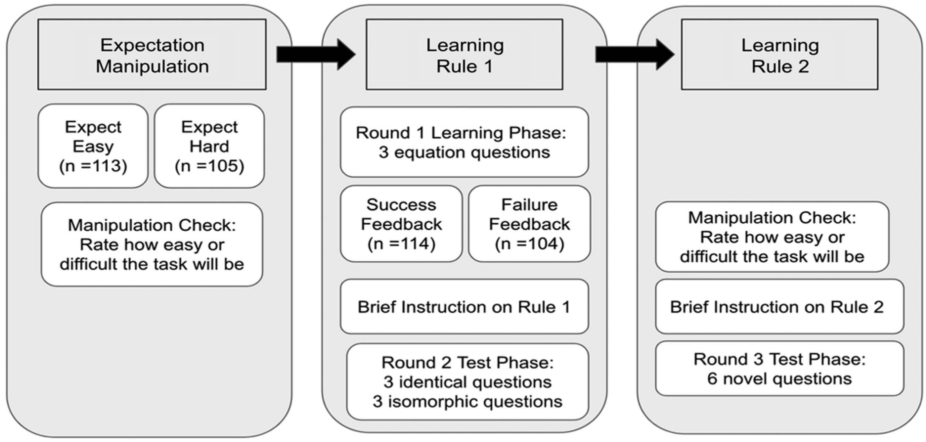
The materials and procedures were fairly similar to Study 2, but with some larger changes to accommodate the rule-based nature of the task and the inclusion of a second rule-based task (see Figure 5). As in Study 2, before Round 1 started, we manipulated participants' expectations about the task (see Figure 3), with some participants being told the task was extremely easy and some being told the task was extremely hard. A manipulation check was then administered in which participants were asked to rate how easy or difficult they expected the task to be on a 5-point Likert scale. Then, participants proceeded to learn about Rule 1, which consisted of a Round 1 learning phase and a Round 2 test phase.

Rule 1. As in previous experiments, participants completed three questions during the Round 1 learning phase. For these questions, participants had to select the symbol that correctly satisfied the equation. However, because the equations were rule-based, the correct response was predetermined based on the rules. Thus, the feedback after each question indicated whether the participants' response was correct or incorrect according to those rules. This procedure meant participants sometimes received a mix of success and failure feedback. In order to create a binary split of feedback valence groups like the previous experiments, participants received an overall feedback message at the end of Round 1 that indicated success or failure. Participants who solved zero to one questions correctly received failure feedback (i.e., "You solved most/all the equations INCORRECTLY!"), and participants who solved two to three questions correctly received success feedback (i.e., "You solved most/all the equations CORRECTLY!"). At this point, participants were also reminded of the task expectations based on their condition (e.g., "Remember that these equations are extremely easy/difficult to solve.") and they were told whether their performance aligned or misaligned with those manipulated expectations. For example, if a participant was told the task was easy and they received success feedback, they were told that their performance aligned with that of most people who took the test. Thus, at the end of Round 1, all participants had answered three rule-based equation questions and could be categorized into one of four groups based on expectations and feedback valence: easy-success, easy-failure, hard-success, or hard-failure.

After Round 1, all participants were given brief instructions about the rule and how to apply the rule to solve the equations (see Figure 6). Then, participants completed Round 2 (the test phase) to assess how well they learned the rule. The first three questions were the same as those asked in Round 1 (identical items). We also included three additional questions that included novel symbols but followed the same rule (isomorphic items). The questions within each of the two item types were randomly ordered.

Figure 5

Schematic Depiction of the Procedures in Study 3



Rule 2. The previous experiments ended immediately after Round 2. However, in this study, the participants were informed that they were going to learn a new rule with a new set of equations. Our goal was to see if preexisting expectations and preexisting experience with success or failure feedback influenced their ability to learn a new set of information. The participants were asked to rate how easy or difficult they expected this second rule to be on a 5-point Likert scale. All participants were then given brief instructions about the new rule and how to apply the rule to solve a set of equations (see Figure 6). Then, they completed a test phase with six questions (Round 3) to assess how well they learned this new rule.

Study 3a: Results

We first report on several manipulation checks to ensure our manipulations worked as intended. We then report on the learning outcomes for Rule 1 based on the questions answered correctly in the Round 2 test phase. Finally, we report on the learning outcomes for Rule 2 based on the questions answered correctly in the Round 3 test phase.

Manipulation Check

We tested whether our expectation manipulation worked by examining participants' expectation ratings out of five, with higher scores representing more difficulty. The manipulation was successful, as participants randomly assigned to the expect-hard condition expected the task to be more difficult (M = 3.94, SD = 0.80) than those randomly assigned to the expect-easy condition (M = 2.30, SD = 0.85), $F(1, 214) = 185.55$, $p < .01$, $\eta^2 = .46$. Given that the expectation manipulation occurred before any feedback was provided, there was no main effect of feedback valence,[5] $F(1, 214) = .12$, $p = .72$, $\eta^2 = .00$, or valence by expectation interaction, $F(1, 214) = .01$, $p = .90$, $\eta^2 = .00$.

We also tested whether performance during the Round 1 learning phase, which determined whether they received success or failure feedback, was at chance. These Round 1 questions had two options (one correct, one incorrect), and participants had not yet learned the rule. Thus, we expected both options to look plausible and to be selected with equal proportions. As expected, one-sampled t tests revealed that both randomized groups (expect-easy and expect-hard) performed at chance level in Round 1 (expect-easy: M = 55%, SD = 35%, p = .12, d = .14; expect-hard: M = 47%, SD = 32%, p = .51, d = −.06).

Rule 1. We analyzed the two types of Round 2 test items (identical and isomorphic) separately. These were given after participants had received feedback in Round 1 and after they had received instruction on the rules. On the identical test items, there was no main effect of expectation condition, $F(1, 214) = .77$, $p = .37$, $\eta^2 = .00$, or feedback valence, $F(1, 214) = .52$, $p = .46$, $\eta^2 = .00$. However, there was an interaction effect of the two, $F(1, 214) = 5.89$, $p = .01$, $\eta^2 = .03$. Exploratory follow-up tests (Bonferroni-adjusted α level = .025 per test) showed that when participants expect the task to be easy, success feedback (M = 92%, SD = 21%) resulted in descriptively higher, but statistically nonsignificant, learning than failure feedback (M = 88%, SD = 27%), $t(82) = .95$, p = .35, d = .19. But, when participants expected the task to be hard, success feedback (M = 88%, SD = 23%) resulted in significantly lower learning than failure feedback (M = 97%, SD = 11%), $t(65) = −2.58$, p = .01, d = −.53.

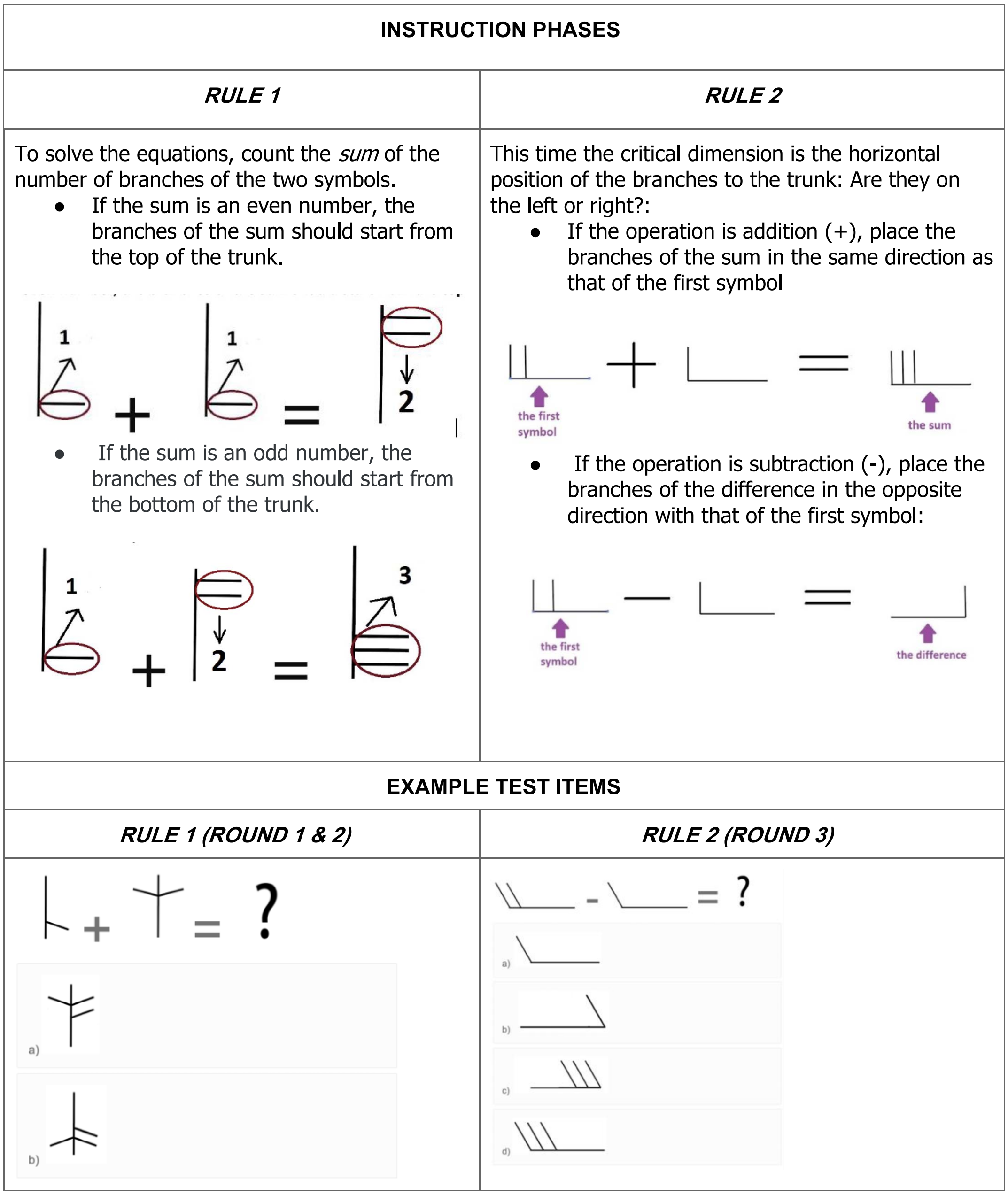
On the isomorphic items, there was not a significant main effect of task-expectation condition, $F(1, 214) = .02$, $p = .88$, $\eta^2 = .00$; feedback valence, $F(1, 214) = .10$, $p = .75$, $\eta^2 = .00$; or the interaction of the two, $F(1, 214) = .04$, $p = .85$, $\eta^2 = .00$.

Rule 2. After Round 1 (learning phase) and Round 2 (test phase), participants were told they were going to learn a second rule. Their expectations about this rule depended on their experience. For predicting their expectation ratings, there was a main effect of expectation condition, $F(1, 214) = 15.32$, $p < .01$, $\eta^2 = .06$, as participants in the expect-hard condition (M = 3.29, SD = 1.05) expected the second rule to be more difficult than those in the

[5] For each statistical test in which we report *feedback valence* as a factor, we also ran a secondary analysis in which we replaced feedback valence with *objective scores (at Round 1)* treated as a continuous variable. All patterns were consistent across the two types of analyses. See the descriptive results of main tests sorted by objective scores and expectation condition at https://osf.io/s9fj5.

Figure 6

Study 2b Rule 1 and Rule 2 With Example Test Items



Note. See the online article for the color version of this figure.

expect-easy condition (M = 2.72, SD = 0.93). There was also a main effect of feedback valence, $F(1, 214) = 5.56$, $p = .01$, $\eta^2 = .02$, as participants who received failure feedback in Round 1 (M = 3.20, SD = 1.06) expected the task to be more difficult than those who received success feedback in Round 1 (M = 2.81, SD = 0.97). There was no interaction, $F(1, 214) = 0.01$, $p = .96$, $\eta^2 = .00$.

In terms of learning the second rule, performance differed by condition. On the Round 3 test items, there was a main effect of feedback valence, $F(1, 214) = 7.50$, $p < .01$, $\eta^2 = .03$. In contrast to the previous studies, those who received failure feedback (M = 78%, SD = 27%) performed better than those who received success feedback (M = 67%, SD = 29%). There was no main effect of expectation condition, $F(1, 214) = .67$, $p = .41$, $\eta^2 = .00$, or an interaction, $F(1, 214) = .00$, $p = .98$, $\eta^2 = .00$.[6]

[6] Finally, we ran a preregistered test on perceived task difficulty of the overall experiment. No factors or their interactions were significant (Fs < 1, $p > .5$)

Study 3b: Method

In Study 3b, we replicated Study 3a with one minor modification. We counterbalanced the order of identical and isomorphic questions in Round 2 to account for potential order effects.

Participants

To determine whether the significant results observed in Study 3a would replicate, we set our effect size of interest at the obtained effect sizes on the significant findings of 3a—specifically the results from Rounds 2 and Round 3. A power analysis using G*Power3 (Faul et al., 2007) indicated that a sample size of 256 would provide a power of $1 - \beta = .80$ to detect a small effect size (partial $\eta^2 = .03$), with a Type 1 error probability (α) of .05.

We had 267 participants initiate the study. Seven were excluded because they started the experiment twice and saw both manipulations of task difficulty. Fifteen were excluded because the data was incomplete (three from the hard condition, two from the easy condition, and remaining without any assignment to conditions). The final number of participants remained at 245. The participants were from the same large university in the Midwestern United States and received credit in their psychology course.

Their average age was 19 years ($SD = 2$). Most students reported their gender as female (73%), followed by male (26%), transgender (one participant), and one participant preferred not to state. Most reported their ethnicity as White (76%), followed by Asian or Asian American (8%), Black or African American (6%), Hispanic or Latino (6%), and Other (4%). Their year in college was freshman (54%), followed by sophomore (29%), junior (11%), and senior (6%).

Procedure

Procedures were identical to Study 3a with only one difference: For Rule 1—Round 2, we counterbalanced the order of the block of identical questions and isomorphic questions.

Study 3b: Results

The manipulation was successful, as participants randomly assigned to the expect-hard condition expected the task to be more difficult ($M = 4.00$, $SD = 0.85$) than those randomly assigned to the expect-easy condition ($M = 2.42$, $SD = 0.88$), $F(1, 241) = 197.89$, $p < .01$, $\eta^2 = .45$. There was no main effect of feedback valence, $F(1, 241) = .07$, $p = .79$, $\eta^2 = .00$, or valence by expectation interaction, $F(1, 241) = .74$, $p = .39$, $\eta^2 = .00$.

We tested whether performance during the Round 1 learning phase, which determined whether they received success or failure feedback, was at chance. One-sampled t tests revealed that the expect-easy group performed significantly better than the chance level ($M = 55\%$, $SD = 30\%$, $p = .04$, $d = .17$), but the expect-hard group's performance was not significantly different than the chance level ($M = 51\%$, $SD = 30\%$, $p = .75$, $d = .03$). Thus, there is a possibility that, for the expect-easy group, the feedback the participants received from one question helped them to figure out the rule and solve the following questions to a small degree.

Rule 1

On the identical test items, there was a main effect of expectation condition, expect-easy: $M = 90.12\%$, $SD = 19.56\%$; expect-hard: $M = 83.33\%$, $SD = 28.10\%$; $F(1, 241) = 5.15$, $p = .02$, $\eta^2 = .02$. There was no main effect of feedback valence, $F(1, 241) = .52$, $p = .42$, $\eta^2 = .00$, and no interaction effect, $F(1, 241) = .14$, $p = .70$, $\eta^2 = .00$.

On the isomorphic items, again, the only significant effect was expectation condition; expectation: $F(1, 241) = 8.01$, $p < .01$, $\eta^2 = .03$; feedback valence: $F(1, 241) = 1.32$, $p = .25$, $\eta^2 = .00$; the interaction: $F(1, 241) = .99$, $p = .31$, $\eta^2 = .00$.

Thus, the interaction effect between feedback valence and expectations observed in Experiment 3a's Round 1 identical items was not replicated in this experiment. To investigate whether timing effects might explain this discrepancy, we conducted an exploratory analysis, collapsing the identical and isomorphic items and analyzing the first and last three items separately. In the analysis of the first half of the items, a significant main effect of expectation was observed, expect-easy: $M = 90.86\%$, $SD = 20.53\%$; expect-hard: $M = 82.72\%$, $SD = 27.36\%$; $F(1, 241) = 7.35$, $p = .007$, $\eta^2 = .02$. However, there was no significant main effect of feedback valence, $F(1, 241) = .55$, $p = .45$, $\eta^2 = .00$, nor an interaction effect, $F(1, 241) = .00$, $p = .95$, $\eta^2 = .00$. A similar pattern emerged for the second half of the items, with a main effect of expectation, expect-easy: $M = 89.87\%$, $SD = 21.65\%$; expect-hard: $M = 82.42\%$, $SD = 27.72\%$; $F(1, 241) = 5.78$, $p = .001$, $\eta^2 = .02$, but no main effect of feedback valence, $F(1, 241) = 1.46$, $p = .22$, $\eta^2 = .00$, and no interaction effect, $F(1, 241) = 1.70$, $p = .19$, $\eta^2 = .00$. These results suggest that the discrepancy between the two experiments cannot be attributed to timing effects.

Rule 2

After Round 1 (learning phase) and Round 2 (test phase), participants were told they were going to learn a second rule. Their expectations about this rule depended on their experience. Participants in the expect-easy group ($M = 2.82$, $SD = .98$) expected the second rule to be easier than those in the expect-hard group ($M = 3.60$, $SD = 1.03$), $F(1, 241) = 33.1$, $p < .001$, $\eta^2 = .12$. There was no main effect of feedback valence, $F(1, 241) = 1.39$, $p = .23$, $\eta^2 = .00$, and no interaction effect, $F(1, 241) = 1.26$, $p = .26$, $\eta^2 = .00$.

In terms of learning the second rule, performance differed by condition. On the Round 3 test items, those who received failure feedback ($M = 70.05\%$, $SD = 28.70\%$) descriptively performed better than those who received success feedback ($M = 64.77\%$, $SD = 32.66\%$), but this difference was not statistically significant, $F(1, 241) = 3.09$, $p = .08$, $\eta^2 = .01$. There was also no main significant effect of expectation condition, $F(1, 241) = 3.15$, $p = .07$, $\eta^2 = .01$. However, there was a significant interaction effect of the two, $F(1, 241) = 5.07$, $p = .02$, $\eta^2 = .02$.

Follow-up tests (Bonferroni-adjusted $\alpha = .025$) revealed that, for the expect-hard condition, failure feedback ($M = 70.97\%$, $SD = 28.36\%$) resulted in significantly better performance than success feedback ($M = 55.12\%$, $SD = 33.90\%$), $t(99.84) = 2.64$, $p < .01$, $d = .5$, which is consistent with the results from Study 3a. For the expect-easy condition, there was no significant difference between failure feedback (M = 69.09%, SD = 29.29%) and success feedback (M = 71.04%, SD = 30.43%) conditions, $t(119.07) = 0.37$, p = .70, d = .07.[7]

Study 3: Discussion

Study 3 investigated the influence of initial failure and success feedback on subsequent tasks. The findings from two experiments suggest that experiencing failure during the learning process does not have a negative effect on learning from subsequent instruction. Furthermore, initial failure had latent advantages in some cases, which became apparent when learning was assessed using a novel task.

For the first rule-based task (learned in Round 1, tested in Round 2), the learning outcomes were typically high across all conditions, and we did not observe consistent condition effects across the two experiments. In Study 3b, we found a main advantage in setting the task expectations to easy for identical items, which is consistent with the findings of Fyfe and Brown (2020), who found that setting expectations for success helped students benefit from later feedback and instruction in mathematics problem-solving. On the other hand, in Study 3a, there was an interaction effect such that failure feedback resulted in better learning than success feedback, but only in the expect-hard condition. Given that these specific results varied across the experiments, it is difficult to draw firm conclusions about how the two factors interact. However, it is worth pointing out that, in Study 3, we switched to a rule-based task with objectively correct answers, and we did not detect any reliable benefits of success feedback relative to failure feedback on any of the outcomes, suggesting that the benefits of success feedback reported in EW-F may not generalize to this context (see Figure 1).

Moreover, the results of learning the second rule-based task were more consistent across the two experiments. Participants who received failure feedback for the first rule outperformed those who received success feedback. This pattern appeared across both expectation manipulations in Study 3a and in the expect-hard condition in Study 3b. We speculate that the benefits of failure feedback did not emerge in the expect-easy condition in Study 3b because it appeared the expect-easy manipulation was particularly salient in that study. Throughout Study 3b, the expect-easy condition uniquely influenced aspects of performance (e.g., the three initial feedback items in Round 1 and the test items in Round 2) in ways that were not present in Study 3a.

We speculate that success feedback was less effective than failure feedback (generally in Study 3a and in the expect-hard condition in Study 3b) because of the illusion of competence it may have induced. In Study 3a, participants who received success feedback may have thought, "I am competent in this task. This was easy for me and I should easily perform well in the rest." Similarly, in Study 3b, the reasoning might have been, "I am successful in this task, even though most people find it very difficult. I should easily perform well in the rest." However, in reality, the tasks for the second rule demanded more cognitive resources than the first one (e.g., a greater number of mental calculations). The illusion of competence at the outset of the second task might have caused participants to exert less effort in the elimination of the distractor options and, in turn, lower performance.

Previous research has similarly emphasized the role of *arousal*, which refers to a state of being alert and activated, and increased activation in the sympathetic nervous system (Arnsten, 2009; Hoogerheide et al., 2019; Wagner et al., 2024). Accordingly, the degree of arousal may vary under different instructional conditions. For example, it has been argued that instructional conditions fostering a sense of competence could result in lower arousal levels and reduced task engagement. Conversely, initial challenges might induce higher levels of arousal, making the task demands salient, leading to increased attentiveness and engagement for the later learning opportunities (Wagner et al., 2024).

Taken together, Study 3 suggests that the tune-out reaction caused by failure feedback does not only vanish when learning from subsequent tasks, but might even lead to more tuning-in than success feedback under some contexts.

General Discussion

The present study replicated and extended the previous work by Eskreis-Winkler and Fishbach (2019) which suggested that people tune out from the task when they receive failure feedback. Our extension initially included the variables of feedback focus (Study 1) and expectations of task difficulty (Study 2). In these studies, learning required memorizing the correct answer provided by the feedback, and the participants were not given any other instructions to learn from, as in the original study by Eskreis-Winkler and Fishbach (2019). In these studies, we replicated the detrimental effects of failure feedback. In Study 3, we turned to a new question: How does failure feedback prepare people to learn from subsequent learning opportunities? In this study, the participants were given feedback on their responses, and then, they were given instructions on how to solve the problems. In this case, learning required understanding the rule and applying it to different examples. We found that the impact of failure feedback on learning was neutral, and sometimes even beneficial relative to success feedback, for the second task structure in Study 3.

Overall, the current research reveals conditions when the detrimental effect of failure feedback on learning is replicated (Study 1 and Study 2), when neither success nor failure feedback has a significant influence on learning (Study 3, the first task), and when people learn better from failure feedback rather than success feedback (Study 3, the second task).

When the task required learners to memorize the meaning of symbols and the outcome measure captured learning *during* the feedback (Study 1 and 2), we replicated the result that people learn less from failure feedback than success feedback. These results were robust to changes in the focus of the feedback message. That is, focusing on the learner (e.g., your answer) versus the task (e.g., the answer) did not influence learning from feedback, and across both variations, failure feedback was worse than success feedback. Also, changing the task expectations (to be easy or hard) did not exert strong influences on learning and generally showed better learning from success feedback. Thus, the outcomes across our first four preregistered experiments are consistent with the hypothesis that failure feedback leads to tune-out reaction *at the moment* of feedback (Eskreis-Winkler & Fishbach, 2019). The previous replication by Keith et al. (2022) found that the negative effects of failure

[7] Finally, on the perceived difficulty of overall experiment, there was a significant main effect of task expectation: expect-easy—M = 2.59, SD = 1.05; expect-hard—M = 2.95, SD = 1.20; $F(1, 241) = 5.56$, $p = .001$, $\eta^2 = .02$.

feedback lessened, but did not vanish, when performance incentives were framed as potential losses rather than gains of learning. Taken together with our findings, it remains to be shown whether task manipulations could make learning from failure as effective as learning from success at the moment of feedback.

While the original effect has been demonstrated as robust, a further question is whether this effect is confined to narrow conditions or if it can be generalized to more typical learning environments, such as schools or workplaces. There is some evidence to suggest that the effect may extrapolate to real-life scenarios. One of EW-F's experiments included a more authentic task where telemarketers took a quiz with facts relevant to their jobs, receiving feedback on either correct or incorrect answers randomly. In addition, observational studies in natural settings, such as clinical practice (Eva et al., 2012; Sargeant et al., 2009; Schartel, 2012) and classroom learning (Fong & Schallert, 2023; Lui & Andrade, 2022), have noted adverse emotional reactions to failure feedback, leading to its reduced reception. However, in these studies, failure feedback was potentially confounded with its natural correlates, such as the informational value of the feedback, interpersonal differences between the participants, and the computational demands of the task. A future challenge for experimental psychology is devising more realistic task scenarios while also separating the feedback valence from such natural correlates. Moreover, there is also ample evidence in the opposite direction: It has been shown that in many cases of factual learning, generating an incorrect response followed by feedback significantly enhances memory for the correct information (for a review, see Metcalfe, 2017). This suggests that detrimental emotional reactions evoked by failure feedback are not universal.

In Study 3, our focus shifted from failure feedback's immediate influence to its impact on subsequent learning opportunities. The task demand also changed in that learners needed to understand the underlying rule and apply it to different examples, as opposed to memorizing symbols. This approach aimed to more closely model real-world scenarios in which people learn a set of rules, apply them to different items, and encounter multiple opportunities for learning after an initial attempt. In this context, we found no evidence that failure feedback was detrimental to learning. These results are consistent with the finding that tune-out effects of failure feedback diminish when the correct answer is provided after feedback (Keith et al., 2022). We also found a novel effect of failure feedback on learning: Failure feedback on the initial task resulted in even better learning from a second, novel task relative to success feedback, especially when the task was expected to be difficult. That is, when given the opportunity to learn a second rule (unrelated to the first rule), those who had previously experienced failure learned more than those who had previously experienced success. This was true regardless of the task expectations in Study 3a, but only in the expect-hard condition in Study 3b.

This latent benefit of failure feedback, as revealed at a later task, is consistent with paradigms such as "productive failure" (Kapur & Bielaczyc, 2012) and "preparation for future learning" (Schwartz & Martin, 2004), which show that problem-solving activities where learners initially fail lead them to learn more from subsequent instruction as opposed to the instructional sequences where such failure is not experienced. An important distinction, however, is the psychological mechanisms we proposed in the current work. These previous works attribute the beneficial effects of initial failure to cognitive processes such as activation of prior knowledge during problem-solving and the ease with which this process helps learners identify critical features of the task during subsequent instruction. Note that this mechanism cannot explain the present study's results. Here, the second rule was entirely independent of the first rule, with only a superficial perceptual resemblance, such as the appearance of symbols and the equations in both tasks. Thus, the present study provides novel evidence that initial failure can be beneficial even when subsequent tasks are structurally unrelated, potentially due to motivational advantages gained from experiencing failure, such as increased arousal and effort. In more natural settings, such as organizational environments, it has also been found that success can eventually decrease attention for future learning, while failure can increase it, disrupting automatic processing and enforcing more effortful information processing (for a review, see, Frese & Keith, 2015). Taken together, these results suggest it may be important to monitor one's performance during learning continuously.

Considering the results from the two task structures, a critical implication from the present study is that the emotional and motivational processes that influence learning *during* failure feedback may be distinguishable from the processes that influence learning *after* failure feedback. This implies that time of learning may be an important factor while considering the tune-out reactions to failure feedback. Based on these initial findings, future research should extend the current paradigm of research by introducing a more systematic temporal variation to learning tasks. It should also be noted that feedback engagement in real life is not necessarily a one-time process, but it is likely to be iterative. Therefore, another route to understanding the role of temporal dynamics in reactions to failure feedback may be providing the opportunity to view feedback at several points.

From a broader theoretical perspective, the conceptual *Model of Motivated Feedback Disengagement* (Grundmann et al., 2021) highlights the temporal dynamics of feedback processing which offers novel insights into the emotional regulation of failure feedback. It suggests that people use several engagement and disengagement strategies when faced with failure feedback based on their performance goals. Accordingly, even though people can initially disengage from the feedback to meet their hedonic goals (e.g., feeling good at that moment), the salience of the hedonic goal decreases in time once it is satisfied. After this decrease, improvement goals become more salient, which motivates feedback engagement again. An important goal for future research should be empirically testing the effective ways of reengaging learners with failure feedback to help meet their goals.

Context

The findings from Eskreis-Winkler and Fishbach (2019) have challenged the long-held assumption that feedback is most helpful for correcting errors across a series of experiments. While these results were surprising to us, we noticed that education researchers have observed similar emotional reactions to failure feedback in school and work settings, such as anxiety, sadness, shame, frustration, and rejection which hinder learners' engagement with feedback (Eva et al., 2012; Fong & Schallert, 2023; Lui & Andrade, 2022; Merrick & Fyfe, 2023). These were primarily studies conducted in authentic learning environments without experimental manipulations. As psychologists interested in learning and cognition, we were inspired by the converging evidence both from these controlled laboratory experiments and from the ethnographic educational research to further investigate causal factors involved in learning from failure feedback and identify strategies to optimize engagement with it.

Constraints on Generality

All our experiments were conducted online with two groups: undergraduate students and MTurk workers, all of whom were participants from the U.S. population. We expect the results to be replicable with different adult populations within the United States. However, we do not have any evidence to suggest that our findings would extend to other cultures, for example, where self-critical views may be particularly motivating for individuals (Heine et al., 2001), or to in-person settings with a human feedback provider, where the influence of interpersonal factors may be more salient (Kluger & Adler, 1993).

References

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology*, *62*(2), 148–156. https://doi.org/10.1037/h0030766

Arnsten, A. F. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, *10*(6), 410–422. https://doi.org/10.1038/nrn2648

Ashford, S. J., & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior and Human Performance*, *32*(3), 370–398. https://doi.org/10.1016/0030-5073(83)90156-3

Butler, A. C., & Woodward, N. R. (2018). Toward consilience in the use of task-level feedback to promote learning. In K. D. Federmeier (Ed.), *Psychology of learning and motivation* (Vol. 69, pp. 1–38). Elsevier Academic Press. https://doi.org/10.1016/bs.plm.2018.09.001

DePasque, S., & Tricomi, E. (2015). Effects of intrinsic motivation on feedback processing during learning. *NeuroImage*, *119*, 175–186. https://doi.org/10.1016/j.neuroimage.2015.06.046

Eskreis-Winkler, L., & Fishbach, A. (2019). Not learning from failure—The greatest failure of all. *Psychological Science*, *30*(12), 1733–1744. https://doi.org/10.1177/0956797619881133

Eva, K. W., Armson, H., Holmboe, E., Lockyer, J., Loney, E., Mann, K., & Sargeant, J. (2012). Factors influencing responsiveness to feedback: On the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education*, *17*(1), 15–26. https://doi.org/10.1007/s10459-011-9290-7

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fong, C. J., & Schallert, D. L. (2023). "Feedback to the future": Advancing motivational and emotional perspectives in feedback research. *Educational Psychologist*, *58*(3), 146–161. https://doi.org/10.1080/00461520.2022.2134135

Frese, M., & Keith, N. (2015). Action errors, error management, and learning in organizations. *Annual Review of Psychology*, *66*(1), 661–687. https://doi.org/10.1146/annurev-psych-010814-015205

Fyfe, E. R., & Brown, S. A. (2020). This is easy, you can do it! Feedback during mathematics problem solving is more beneficial when students expect to succeed. *Instructional Science*, *48*(1), 23–44. https://doi.org/10.1007/s11251-019-09501-5

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology*, *108*(1), 82–97. https://doi.org/10.1037/edu0000053

Gilman, D. A. (1969). Comparison of several feedback methods for correcting errors by computer-assisted instruction. *Journal of Educational Psychology*, *60*(6), 503–508. https://doi.org/10.1037/h0028501

Gok, S., & Fyfe, E. (2024, June 4). *Feedback on failure*. https://osf.io/j85ep/

Grundmann, F., Scheibe, S., & Epstude, K. (2021). When ignoring negative feedback is functional: Presenting a model of motivated feedback disengagement. *Current Directions in Psychological Science*, *30*(1), 3–10. https://doi.org/10.1177/0963721420969386

Guthrie, J. T. (1971). Feedback and sentence learning. *Journal of Verbal Learning and Verbal Behavior*, *10*(1), 23–28. https://doi.org/10.1016/S0022-5371(71)80088-9

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Heine, S. J., Lehman, D. R., Ide, E., Leung, C., Kitayama, S., Takata, T., & Matsumoto, H. (2001). Divergent consequences of success and failure in Japan and North America: An investigation of self-improving motivations and malleable selves. *Journal of Personality and Social Psychology*, *81*(4), 599–615. https://doi.org/10.1037/0022-3514.81.4.599

Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & van Gog, T. (2019). Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology*, *111*(1), 45–56. https://doi.org/10.1037/edu0000272

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Keith, N., Horvath, D., Klamar, A., & Frese, M. (2022). Failure to learn from failure is mitigated by loss-framing and corrective feedback: A replication and test of the boundary conditions of the tune-out effect. *Journal of Experimental Psychology: General*, *151*(8), e19–e25. https://doi.org/10.1037/xge0001170

Kluger, A. N., & Adler, S. (1993). Person- versus computer-mediated feedback. *Computers in Human Behavior*, *9*(1), 1–16. https://doi.org/10.1016/0747-5632(93)90017-M

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, *7*(3), 67–72. https://doi.org/10.1111/1467-8721.ep10772989

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, *47*(2), 211–232. https://doi.org/10.3102/00346543047002211

Lui, A. M., & Andrade, H. L. (2022). The next black box of formative assessment: A model of the internal mechanisms of feedback processing. *Frontiers in Education*, *7*, Article 751549. https://doi.org/10.3389/feduc.2022.751549

Mandouit, L., & Hattie, J. (2023). Revisiting "The Power of Feedback" from the perspective of the learner. *Learning and Instruction*, *84*, Article 101718. https://doi.org/10.1016/j.learninstruc.2022.101718

Mangels, J. A., Good, C., Whiteman, R. C., Maniscalco, B., & Dweck, C. S. (2012). Emotion blocks the path to learning under stereotype threat. *Social Cognitive and Affective Neuroscience*, *7*(2), 230–241. https://doi.org/10.1093/scan/nsq100

Merrick, M., & Fyfe, E. R. (2023). Feelings on feedback: Children's emotional responses during mathematics problem solving. *Contemporary Educational Psychology*, *74*, Article 102209. https://doi.org/10.1016/j.cedpsych.2023.102209

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*(1), 465–489. https://doi.org/10.1146/annurev-psych-010416-044022

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology*, *96*(1), 165–173. https://doi.org/10.1037/0022-0663.96.1.165

Norem, J. K., & Cantor, N. (1986). Defensive pessimism: Harnessing anxiety as motivation. *Journal of Personality and Social Psychology*, *51*(6), 1208–1217. https://doi.org/10.1037/0022-3514.51.6.1208

Sargeant, J. M., Mann, K. V., van der Vleuten, C. P., & Metsemakers, J. F. (2009). Reflection: A link between receiving and using assessment feedback. *Advances in Health Sciences Education*, *14*(3), 399–410. https://doi.org/10.1007/s10459-008-9124-4

Schartel, S. A. (2012). Giving feedback—An integral part of education. *Best Practice & Research: Clinical Anaesthesiology*, *26*(1), 77–87. https://doi.org/10.1016/j.bpa.2012.02.003

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759–775. https://doi.org/10.1037/a0025140

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129–184. https://doi.org/10.1207/s1532690xci2202_1

Son, J. Y., & Goldstone, R. L. (2009). Contextualization in perspective. *Cognition and Instruction*, *27*(1), 51–89. https://doi.org/10.1080/07370000802584539

Wagner, S., Sibley, L., Weiler, D., Burde, J. P., Scheiter, K., & Lachner, A. (2024). The more, the better? Learning with feedback and instruction. *Learning and Instruction*, *89*, Article 101844. https://doi.org/10.1016/j.learninstruc.2023.101844

Received May 15, 2023
Revision received June 4, 2024
Accepted June 8, 2024 ■