Manuscript accepted for publication in *Journal of Experimental Psychology: General*.

Failure to Learn From Failure is Mitigated by Loss-Framing and by Corrective Feedback: A Replication and Test of Boundary Conditions of the Tune-out Effect

Nina Keith[1], Dorothee Horvath[1], Alexander Klamar[2], & Michael Frese[3, 4, 5]

[1]Technical University of Darmstadt

[2]Helmut-Schmidt-Universität/ Universität der Bundeswehr Hamburg

[3]Leuphana University Lüneburg

[4]Asia School of Business (in collaboration with MIT Sloan Management)

[5]National University of Singapore

Author Note

The pre-registrations of the experiments can be found at https://aspredicted.org/de9ck.pdf (Study 1) and at https://aspredicted.org/2nx2v.pdf (Study 2). The data of both experiments can be found at https://researchbox.org/178.

This research was supported by grants from the German Research Foundation (Deutsche Forschungsgemeinschaft, grants no. FR 638/38-1 and KE 1377/4-1). We would like to thank Hannah Habicht and Verena Schieck for data collection.

Correspondence concerning this article should be addressed to Nina Keith, Organizational and Business Psychology, Technical University of Darmstadt, Alexanderstr. 10, 64283 Darmstadt, Germany. E-mail: keith@psychologie.tu-darmstadt.de

Abstract

Do people learn from failure or do they mentally "tune-out" upon failure feedback, which in turn undermines learning? Recent research (Eskreis-Winkler & Fishbach, 2019) suggests the latter, while research in educational and work settings indicates that failure can lead to more learning than success and error-free performance. We conducted two pre-registered experiments to replicate the tune-out effect and to test two potential boundary conditions (total $N = 520$). The tune-out effect fully replicated in those experimental conditions that represented close replications of the original study, underscoring the reliability of the original effect. However, the effect disappeared when the same monetary incentives for participation were expressed in terms of a loss (i.e., losing money for each wrong answer) rather than a gain (i.e., earning money for each correct answer) (Experiment 1). The effect also disappeared when additional corrective feedback was given (Experiment 2). It seems that switching from gain to loss framing or giving corrective feedback (vs. no corrective feedback) are substantial and meaningful variations of the original paradigm that constitute boundary conditions of the tune-out effect. These results help explain the conflicting findings on learning from failure and suggest that in many applied settings, tuning out upon failure may not be an option.

Keywords: learning from errors, learning from failure, loss aversion, corrective feedback, preregistration

Failure to Learn From Failure is Mitigated by Loss-Framing and by Corrective Feedback: A Replication and Test of Boundary Conditions of the Tune-out Effect

Do people learn from failure? Or does failure lead people to "tune out" and stop paying attention, which undermines learning? Recent research suggests the latter: In a series of experiments, Eskreis-Winkler and Fishbach (2019) found participants who received success feedback to consistently outperform those who received failure feedback. These findings are in line with earlier research on negative motivational consequences of failure (e.g., Soman & Cheema, 2004) and potential undesirable consequences of feedback (e.g., Kluger & DeNisi, 1996). They also extend earlier research by suggesting that these negative consequences occur very early in the motivational process, spurred by an immediate motivational shutdown (tune-out effect) in the face of failure.

In contrast to their findings, there is a vivid stream of applied research that suggests the opposite, namely, that failure and errors facilitate learning more than successes and error-free performance.[1] The idea is that failure grabs people's attention (Vaish et al., 2008; Zakay et al., 2004) and creates a sense of urgency to respond, more so than does success (Audia et al., 2000; Horvath et al., 2021; Madsen & Desai, 2010; Sitkin, 1992). People then engage in mindful deeper-level processing that increases learning (Cannon & Edmondson, 2004; Ivancic & Hesketh 2000). A number of empirical studies from organizational and educational settings support these arguments (Frese & Keith, 2015; Keith & Frese, 2008; Metcalfe, 2017).

How can these seemingly opposing views and findings be reconciled? In two preregistered experiments, we tested two potential boundary conditions that may mitigate the tune-out effect: (1) the framing of incentives and (2) provision of corrective feedback. We propose that psychologically meaningful differences in these two context characteristics contribute to the conflicting findings. Below, we develop these points in more detail.

Framing of Incentives

First, we expected the tune-out effect to appear under a gain-framing (as in the original study) but not under a loss-framing (Hypothesis 1), for the following reasons. The original study incentivized performance by offering monetary rewards for correct test answers. This inadvertently created a gain-framing. People respond differently to gains and losses and are generally loss-aversive (Kahneman & Tversky, 1979). In applied settings, although performance incentives and other gains may be common, failure is typically associated with the threat of loss (e.g., fear of damaging one's reputation or even job loss; Zhao & Olivera, 2006). The threat of loss, more than the prospect of gain, may motivate learning from failure to avoid the anticipated future loss that comes with poor performance. In other words, while failure may generally be aversive and trigger a tune-out reaction, the prospect of loss may be even more aversive and override or at least reduce the first impulse to tune out. To test this possibility, we introduced experimental conditions with a loss-framing of incentives, while keeping the monetary regime exactly the same as in the gain-framing conditions.

Provision of Corrective Feedback

Second, we focused on corrective feedback (i.e., feedback that presents the correct solution). We expected the tune-out effect to appear when no such feedback is given (as in the original study) but not when corrective feedback is given (Hypothesis 2), for the following reasons. With their novel research paradigm, Eskreis-Winkler and Fishbach (2019) elegantly created a situation in which mere outcome feedback (correct/success vs. incorrect/failure) includes the full information on correct answers. Specifically, because it was a binary-choice task, success feedback technically contained the same information on the correct answer as failure feedback. However, to excerpt and learn the correct answer from feedback, participants of the failure condition had to process a negation (i.e., "This answer was *not* correct", which in itself can be a difficult task; Carpenter & Just, 1975), make an inference based on this negation (i.e., "Thus, the other answer must have been the correct one"), and remember the response they had *not* chosen. Participants of the success condition, in contrast, received immediate reinforcement for the response they had just chosen. In applied settings, mere outcome feedback is often accompanied with some informational feedback on what one should have done instead, for example, when a teacher returns a math test to a student, when a supervisor discusses an error with an employee, or when managers discuss a failed project. Indeed, some scholars explicitly state that feedback should include information on correct responses to be effective (Hattie & Timperley, 2009; Metcalfe, 2017). In other words, the failure condition in Eskreis-Winkler and Fishbach's studies may have imposed higher cognitive demands during learning on participants of the failure condition than on those of the success condition and also than in many applied feedback situations. Provision of corrective feedback may reduce these cognitive demands and facilitate learning from failure (Pashler et al., 2005).

Method

Experimental Design

We conducted two pre-registered experiments that used a 2x2 between-subjects design with two levels each. In both experiments, Factor 1 was outcome feedback with the two levels success ("Correct") vs. failure ("Incorrect"), as in the original studies by Eskreis-Winkler and Fishbach (2019, Studies 2a and 2b). Factor 2 reflected our two hypotheses on framing of incentives (tested in Experiment 1, extension of Study 2a) and corrective feedback (tested in Experiment 2, extension of Study 2b). Accordingly, Factor 2 in Experiment 1 was the framing of performance incentives, with the two levels *gain-framing* (as in the original study) and *loss-framing*. Factor 2 in Experiment 2 was provision of corrective feedback, with the two levels *no corrective feedback* (as in the original study) and *corrective feedback*. In other words, of the 4 cells in our 2x2 designs, 2 cells were replications of the original study by Eskreis-Winkler and Fishbach (2019). The other 2 cells were variations of their basic design that we introduced to test our hypotheses.

Participants and Sample Size

We determined the optimal sample size prior to data collection (see pre-registrations). Eskreis-Winkler and Fishbach's focal analyses were *t* tests that compared success vs. failure feedback conditions. The associated effects they found were medium sized (Cohen's *d* of 0.51 to 0.72). Based on these analyses and effect sizes, optimal sample size is 128, according to G*Power (Faul et al., 2007) (medium effect of 0.50, with a type-I error probability of $\alpha = .05$ and a power of $1-\beta = .80$). Because we had 2 factors and 4 cells (rather than 1 factor with 2 cells, as the original study), we doubled the sample of 128 and targeted an overall sample of 256 for both experiments, respectively. We recruited more participants to account for potential data loss. In Experiment 1, our final sample included 260 adult participants (36.2% female, mean age = 35.61 years, *SD* = 10.34), most of who worked full-time or part-time (95.4%). In Experiment 2, too, our final sample included 260 adult participants (40.0% female, mean age = 35.77 years, *SD* = 10.74), most of who worked full-time or part-time (93.1%). We used the same way of recruitment as the original study, namely, Amazon's Mechanical Turk (MTurk). Only individuals who had an MTurk approval rating of 50% or higher were allowed to participate (same as in the original study). Also, as preset in our preregistrations, we employed attention-check items and excluded participants with careless responses (Meade & Craig, 2012). In particular, we included three types of attention checks. The first one asked participants for the incentive scheme; participants chose 1 out of 4 answer choices presented to them (for Study 1, this attention check also constituted a manipulation check because the incentive scheme was an experimental factor in Study 1). Participants who chose a wrong answer were not allowed to continue. The second one involved three nonsensical, bogus items (e.g., "I receive my paycheck from goblins."; Meade & Craig, 2012) that we embedded within additional variables we assessed (see below). We informed participants that there would be such items (including one sample item), the purpose of these items (namely, to screen out participants who do not read instructions and questions carefully), and that they needed to answer these items meaningfully in order to be allowed to continue and to be paid. Finally, the third attention check was the same as the one used in the original studies: In the beginning of the survey and between the learning and test phase, participants were asked to write answers to short question (i.e., to name one's favorite book and music, see below) into an open textbox. Only participants who answered these questions were included in the analyses (cf. Eskreis-Winkler & Fishbach, 2019, p. 1737). In Web-experiments, selective attrition may be a serious issue that can lead to false research conclusions (Zhou & Fishbach, 2016). Because in our experiments, all participants received identical attention checks irrespective of experimental conditions, attrition bias seems unlikely. A flowchart depicting the attrition numbers in experimental groups is depicted in our additional material (https://researchbox.org/178). Attrition did not differ statistically between gain vs. loss conditions in Experiment 1, χ^2 (df = 1) = 0.135, p = .714, or between the corrective feedback vs. no corrective feedback conditions in Experiment 2, χ^2 (df = 1) = 2.394, p = .122.

Participants consented to participation before starting the actual survey. Procedures were in line with ethical standards of the German Psychological Society (DGPs). Formal approval by an ethics committee was not required.

Material and Procedure

The material and task were identical to those used by Eskreis-Winkler and Fishbach (2019), namely, a script task (original material available at https://osf.io/q7wkv/?view_only=61b2d03f3a0f460dad6de333d08a0459; material of the present experiments available at https://researchbox.org/178).

Round 1 was the learning phase and consisted of three questions that were presented to participants in a randomized order. "Each question asked participants to guess which of two symbols had a specific meaning in an invented language"(Eskreis-Winkler & Fishbach, 2019, p. 1737), for example: "Which of the following characters in an ancient script represents an animal?  or ". After participants had chosen an answer, the manipulation of Factor 1 (success vs. failure feedback) took place. Participants in the *success feedback* conditions received success feedback after each question ("You answered this question correct!", Eskreis-Winkler & Fishbach, 2019, p. 1741). Participants in the *failure feedback* conditions received failure feedback after each question ("You answered this question incorrect!", p. 1741). Participants received the same feedback for all three questions, that is, only success or only failure feedback.

Round 2 was the test phase that measured participants' learning. As in Round 1, participants responded to three questions that were presented in a randomized order. The three questions "paralleled each of the initial questions (Round 1) but were phrased in the reverse" (Eskreis-Winkler & Fishbach, 2019, p. 1737). For example, the question "Which of the following characters in an ancient script represents a bird?" in Round 1, was changed in Round 2 to "Which of the following characters represents a non-living, stationary object. Both the ordering of the test-phase questions and, within the questions, the ordering of the answer choices were randomized. Participants were informed from the beginning that there would be two rounds and that Round 2 would be the test phase. This information was included in the welcome text and stressed once again before Round 1 started.

Between Round 1 and Round 2, participants responded to a distractor question: "Tell us: what is your favorite music to listen to?" (Eskreis-Winkler & Fishbach, 2019, p. 1737). This question also served to screen out participants (i.e., only participants who would provide an answer to this question were included; cf. Eskreis-Winkler & Fishbach, 2019). After Round 2, they completed questionnaires on person characteristics. Finally, participants were informed about their performance in Round 2 of the script task, thanked, and compensated. The compensation was a minimum of $.050 and a maximum of $2.00 in both experiments. (In the original studies, performance incentives varied between $0.10 for each correct test answer in Study 2a and $1.50 in Study 2b, with similar results.)

In Experiment 1, we used the compensation schedule to manipulate Factor 2 (gain-framing vs. loss-framing). Participants in the *gain-framing* conditions were informed that they will be paid a compensation of $0.50 for mere participation and that they would have the chance to gain $0.50 for each question they would answer correctly on the test, which adds to a maximum of an additional $1.50 (i.e., up to $ 2.00) because the test had three questions. Participants in the *loss-framing* conditions were informed that they will be paid a maximum compensation of $2.00. Of this $2.00 credit that they have in the beginning, however, they would lose $0.50 for each question they answered incorrectly on the test, which means they would still be left with $0.50 if they answered all test questions incorrectly. Thus, the payment structure was identical for the gain-framing and the loss-framing conditions, but only the linguistic framing differed. We presented the information on the compensation schedule twice, namely, upon entry to the survey and once again before the learning phase began.

In Experiment 2, compensation was identical for all participants (all received $0.50 for mere participation and an additional $1.50 if they answered the test questions correctly). Here, the manipulation of Factor 2 involved whether participants received additional corrective feedback or not. That is, participants who did not receive this corrective feedback only received the success vs. failure feedback as described above (i.e., "Correct" vs. "Incorrect", as in the original study). In those conditions that included corrective feedback, participants were additionally shown the correct answer before the next question was presented ("The correct answer is …").

Measures

Dependent Variable

The focal dependent variable was learning, measured as the percentage of the test-phase questions that a participant answered correctly (same as in the original study).

Additional Variables

As preset in our preregistration, we assessed individual-difference variables (person characteristics) for exploratory purposes in both experiments. These included *negative affectivity* as a trait (Watson et al., 1988; 10 items, Cronbach's α = .95 in both experiments) and *learning from error* as an attitude towards errors as an opportunity to learn (Rybowiak et al., 1999; 4 items, sample item "Mistakes assist me to improve my work", Cronbach's α = .85 in Experiment 1 and .93 in Experiment 2). We chose these two person characteristics because previous research indicates that they may be associated with learning from failure (e.g., Keith et al., 2020).

Results

Experiment 1: Framing of Incentives

Hypothesis 1 predicted success feedback to lead to better learning (i.e., more correct answers on the test) than failure feedback under a gain-framing, as in the original study, but not under a loss-framing. To replicate Eskreis-Winkler and Fishbach's (2019) statistical analyses as closely as possible, we first ran *t* tests for the gain-framing conditions and the loss-framing conditions separately. In line with Hypothesis 1, success feedback led to better learning than failure feedback in the gain-framing conditions, $t(128) = 6.98$, $p < .001$, Cohen's $d = 1.23$, but not in the loss-framing conditions, $t(128) = 1.48$, $p = .141$, Cohen's $d = 0.26$ (Table 1 and Figure 1, top panel).

Second, we tested whether participants of any condition learned anything at all, that is, whether their test performance exceeded chance level (i.e., 1.5 out of 3 or 50%; cf. Eskreis-Winkler & Fishbach, 2019). In line with the results found in the original study, under a gain-framing, learning exceeded chance level in the success condition, $t(67) = 12.32$, $p < .001$, Cohen's $d = 1.49$, but not in the failure condition, $t(61) = 0.37$, $p = .710$, Cohen's $d = 0.05$. Under a loss-framing, in contrast, learning exceeded chance level both in the success condition, $t(70) = 6.99$, $p < .001$, Cohen's $d = 0.83$, and in the failure condition, $t(58)=4.29$, $p < .001$, Cohen's $d = 0.56$. Again, this pattern is supportive of Hypothesis 1.

Third, because our experiment had a two-factorial design (Factor 1: success vs. failure feedback; Factor 2: gain vs. loss framing), we ran a 2x2 between-subjects ANOVA. In support of Hypothesis 1, the results parallel those of the aforementioned *t* tests in that the two factors interacted, $F(1,256) = 11.18$, $p < .001$, $\eta_p^2 =.04$. Outcome feedback (success vs. failure feedback) had a main effect, $F(1,256) = 31.52$, $p < .001$, $\eta_p^2 =.11$. Framing did not have a main effect, *F*(1,256) = 1.36, *p* = .244, η_p^2 =.01. The results remained stable when including individual-difference variables negative affectivity and attitude towards learning from errors as covariates, interaction *F*(1,254) = 10.59, *p* = .001, η_p^2 =.04.

Experiment 2: Corrective Feedback

Hypothesis 2 predicted success feedback to lead to better learning (i.e., more correct answers on the test) than failure feedback when only outcome feedback (i.e., success vs. failure feedback) is given, as in the original study, but not when additional corrective feedback is given. To replicate Eskreis-Winkler and Fishbach's (2019) statistical analyses as closely as possible, we first ran *t* tests for the conditions without and with corrective feedback separately. In line with Hypothesis 2, success feedback led to better learning than failure feedback when no corrective feedback was given, *t*(128) = 3.99, *p* < .001, Cohen's *d* = 0.71, but not when corrective feedback was given, *t*(128) = 1.75, *p* = .083, Cohen's *d* = 0.31 (Table 1 and Figure 1, bottom panel).

Second, we tested whether participants of any condition learned anything at all, that is, whether their test performance exceeded chance level (i.e., 1.5 of 3 or 50%; cf. Eskreis-Winkler & Fishbach, 2019). In all experimental conditions, learning exceeded chance level: *t*(57)=9.63, *p* < .001, Cohen's *d* = 1.26 for success and no corrective feedback; *t*(71)=2.49, *p* = .015, Cohen's *d* = 0.29 for failure and no corrective feedback; *t*(76)=9.62, *p* < .001, Cohen's *d* = 1.10 for success and corrective feedback; and *t*(52)=5.54, *p* < .001, Cohen's *d* = 0.76 for failure and corrective feedback.

Third, because our experiment had a two-factorial design (Factor 1: success vs. failure feedback; Factor 2: with vs. without additional corrective feedback), we ran a 2x2 between-subjects ANOVA. The results did not fully parallel those of the aforementioned *t* tests in that the two factors did not interact, *F*(1,256) = 3.01, *p* = .084, η_p^2 =.01. Outcome feedback (success vs.

failure feedback) had a main effect, $F(1,256) = 16.86$, $p < .001$, $\eta_p^2 = .06$. Corrective feedback did not have a main effect, $F(1,256) = 1.99$, $p = .159$, $\eta_p^2 = .01$. However, the results changed when we included individual-difference variables negative affectivity and attitude towards learning from errors as covariates: parallel to the *t* tests reported above, and in support of Hypothesis 2, the factors interacted, $F(1,254) = 4.62$, $p = .033$, $\eta_p^2 = .02$.

Discussion

About 90 years ago, Thorndike (1931) observed an asymmetry in learning from positive and negative feedback: "Other things being equal, an announcement of 'Right' strengthens the connection which it follows and belongs to much more than the announcement of 'Wrong' weakens the one which it follows and belongs to" (p. 45). Using a novel paradigm, a recent study (Eskreis-Winkler & Fishbach, 2019) found a similar effect, namely, more learning from success feedback ("satisfiers", in Thorndike's terms) than from failure feedback ("annoyers", in Thorndike's terms). Presumably, people tend to mentally tune out upon failure which, in turn, undermines learning. On the other hand, applied research consistently demonstrated that errors and failure can be effective drivers of learning. In an attempt to reconcile these seemingly contradictory findings, the present research proposed two boundary conditions of the tune-out effect, namely, the provision of corrective feedback and the gain versus loss framing of performance incentives. As predicted, in two preregistered experiments, these two context characteristics mitigated the tune-out effect. In addition, this research included experimental conditions that represented close replications of the original study. In these conditions, the tune-out effect was fully replicated, which underscores the reliability of the original effect.

Our research extends previous findings by identifying two practically important context characteristics that help specify the conditions under which tuning out upon failure may be more or less likely. Specifically, at work and in educational settings, where longer-term commitment, social control, and threat of negative performance consequences are prevalent, simply tuning out and mentally or even physically disengaging after failure may not be a viable option. Also, in an applied context, it is unlikely that people receive mere failure feedback without the opportunity to discuss or reflect upon alternative courses of action (i.e., corrective feedback). The context characteristics we identified not only help explain the contradictory findings. They also have implications as to how learning from failure in organizations can be enhanced, for example, by providing explicit opportunities to discuss errors and failure in a psychologically safe atmosphere (Edmondson & Lei, 2014; Zhao et al., 2018) and by promoting a constructive organizational culture of error management (van Dyck et al., 2005). Such measures may counter the tendency to mentally tune-out upon failure and to avoid reflecting on the potentially insightful failure experience.

That being said, it should be noted that although the two context characteristics did mitigate the effect, they did not reverse it. That is, in the conditions we introduced, participants learned similarly from failure and from success feedback; there was no evidence that they learned more from failure than from success feedback. In all, our results support the idea by Eskreis-Winkler and Fishbach (2019) that when the incentive to learn is not too high, it seems to be easier to learn from success than from failure feedback. Our results extend their findings by demonstrating that this learning advantage of success feedback may diminish when the motivation to learn is higher (e.g., as under loss-framing) or when the task is made easier (i.e., as under provision of corrective feedback). With regard to provision of corrective feedback, it should also be noted that the research question we asked was somewhat different from the original one, in that we asked whether people could learn from corrective feedback after success vs. failure feedback, rather than from success vs. failure feedback per se.

Apart from the provision of feedback and framing of incentives, which the present research identified as buffers of the tune-out effect, there may be a number of additional context characteristics worth studying. In particular, the study by Eskreis-Winkler and Fishbach (2019) and our extension of their research used one particular task, namely, a binary-choice guessing task. This task has two clear-cut advantages. First, it provides an objective performance measure and, second, it keeps the experience of success versus failure fully independent of and unconfounded with participants' previous task knowledge and abilities (i.e., the two most powerful predictors of learning and performance in applied settings). Yet, it remains to be tested whether the observed tune-out effect also generalizes to a wider range of tasks that are more typical in organizational and educational settings. For example, research suggests that there are fundamental differences in learning from actual errors versus mere wrong guesses and from errors made with high versus with low confidence (i.e., the degree to which people believe that their error is correct) (Metcalfe, 2017). In organizational and educational settings, many errors may be of the high-confidence type in that employees, managers, and students are confident that their (erroneous) decisions and courses of actions were correct, as they act to the best of their knowledge. It is plausible to assume that employees' and managers' basic motivation to learn from such errors is higher to begin with, making a motivational shutdown and tuning-out upon failure less likely.

Critics may even argue that the presently used experimental task was a far cry from the kind of failure and errors people experience in applied settings and that the presently used paradigm did not actually study failure at all. Indeed, some scholars consider the feeling of "I should have known better" to be constitutive of an error (Frese & Keith, 2015; Hofmann &

Frese, 2011)—a feeling that was not possible in the presently used mere guessing task, because whether a response was right or wrong was not meaningful but arbitrarily set by the experimental procedure. Also, we cannot be sure whether participants, that is, an anonymous online sample working for some extra payment, really felt a sense of failure at all that would cause them to tune out based on an emotional response. Eskreis-Winkler and Fishbach (2019) present some evidence in support of their tune-out interpretation (specifically, they showed a mediating and a moderating effect of ego threat in Studies 4 and 5, respectively). Still, future research may try and test more directly the role of emotions as explanatory mechanism for the here observed effects.

Context

With a background in applied psychology and management, much of our past research on errors and failure dealt with potentially positive outcomes, particularly in terms of learning from errors and stimulating innovations in individuals and organizations (Frese & Keith, 2015; Keith & Frese, 2008). When we learned about the results by Eskreis-Winkler and Fishbach (2019), who consistently demonstrated better learning from success than from failure across multiple experiments, we were astonished, fascinated, but we also felt challenged. We appreciated their article because it also made it possible to have a serious debate about learning from errors, which is a fascinating and important topic that warrants a discussion across basic and applied fields of psychology. We decided we should replicate their effect independently. But we also had several ideas about how the divergence of findings may be explained and we sought to put them to the test. Finally, we came up with this combination: a replication and variation of the experiments that would test at least some of our ideas.

References

Audia, P. G., Locke, E. A., & Smith, K. G. (2000). The paradox of success: An archival and laboratory study of strategic persistence following radical environmental change. *Academy of Management Journal*, *43*, 837–853. https://doi.org/10.5465/1556413

Cannon, M. D., & Edmondson, A. C. (2005). Failing to learn and learning to fail (intelligently): How great organizations put failure to work to innovate and improve. *Long Range Planning*, *38*(3), 299-319. https://doi.org/10.1016/j.lrp.2005.04.005

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*(1), 45-73. https://psycnet.apa.org/doi/10.1037/h0076248

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%2F0956797619881133

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

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

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior*, 1, 23-43. https://doi.org/10.1146/annurev-orgpsych-031413-091305

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

Hofmann, D. A., & Frese, M. (2011). Errors, error taxonomies, error prevention, and error management: laying the groundwork for discussing errors in organizations. In D. A. Hofmann & M. Frese (Eds.), *Errors in organizations* (pp. 1–44). Taylor & Francis.

Horvath, D., Klamar, A., Keith, N., & Frese, M. (2021). Are all errors created equal? Testing the effect of error characteristics on learning from errors in three countries. *European Journal of Work and Organizational Psychology*, *30*(1) 110-124. https://doi.org/10.1080/1359432X.2020.1839420

Ivancic, K., & Hesketh, B. (2000). Learning from errors in a driving simulation: Effects on driving skill and self-confidence. *Ergonomics*, *43*(12), 1966–1984. https://doi.org/10.1080/00140130050201427

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291. http://www.jstor.org/stable/1914185

Keith, N., & Frese, M. (2008). Effectiveness of error management training: a meta-analysis. *Journal of Applied Psychology*, *93*(1), 59-69. https://doi.org/10.1037/0021-9010.93.1.59

Keith, N., Horvath, D., & Klamar, A. (2020). The more severe the merrier: Severity of error consequences stimulates learning from error. *Journal of Occupational and Organizational Psychology*, *93*, 712-737. https://doi.org/10.1111/joop.12312

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

Madsen, P. M., & Desai, V. (2010). Failing to learn? The effects of failure and success on organizational learning in the global orbital launch vehicle industry. *Academy of Management Journal*, *53*(3), 451-476. https://doi.org/10.5465/amj.2010.51467631

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

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

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When Does Feedback Facilitate Learning of Words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3-8. https://doi.org/10.1037/0278-7393.31.1.3

Rybowiak, V., Garst, H., Frese, M., & Batinic, B. (1999). Error Orientation Questionnaire (EOQ): Reliability, validity, and different language equivalence. *Journal of Organizational Behavior*, *20*(4), 527–547. https://doi.org/10.1002/(SICI)1099-1379(199907)20:4<527::AID-JOB886>3.0.CO;2-G

Sitkin, S. (1992). Learning Through Failure : The Strategy of Small Losses. *Research in Organizational Behavior*, *14*, 231-266.

Soman, D., & Cheema, A. (2004). When goals are counterproductive: The effects of violation of a behavioral goal on subsequent performance. *Journal of Consumer Research*, *31*(1), 52-62. https://doi.org/10.1086/383423

Thorndike, E. L. (1931). *Human learning*. The Century Co. https://doi.org/10.1037/11243-000

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*, 383–403. https://doi.org/10.1037/0033-2909.134.3.383

Van Dyck, C., Frese, M., Baer, M., & Sonnentag, S. (2005). Organizational error management culture and its impact on performance: a two-study replication. *Journal of Applied Psychology*, *90*(6), 1228-1240. https://doi.org/10.1037/0021-9010.90.6.1228

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070. https://psycnet.apa.org/doi/10.1037/0022-3514.54.6.1063

Zakay, D., Ellis, S., & Shevalsky, M. (2004). Outcome value and early warning indications as determinants of willingness to learn from experience. *Experimental Psychology*, *51*, 150–157. https://doi.org/10.1027/1618-3169.51.2.150

Zhao, B., & Olivera, F. (2006). Error reporting in organizations. *Academy of Management Review*, *31*(4), 1012-1030. https://doi.org/10.5465/amr.2006.22528167

Zhao, B., Seifried, J., & Sieweke, J. (2018). Trainers' responses to errors matter in trainees' learning from errors: evidence from two studies. *Journal of Managerial Psychology, 33*(3), 279-296. https://doi.org/10.1108/JMP-10-2017-0364

Zhou, H., & Fishbach, A. (2016). The pitfalls of experimenting on the Web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493-504. https://doi.org/10.1037/pspa0000056

Footnote

[1] Strictly speaking, failure (i.e., an adverse outcome) and errors (i.e., actions that may or may not lead to adverse outcomes depending on the specific circumstances) are not synonymous (Frese & Keith, 2015; Zhao & Olivera, 2006).

Table 1

Comparison of Results from the Test Phases of Experiments 1 and 2, by Condition

Experiment & condition	*N*	Correct answers in failure condition	Correct answers in success condition	Between-group comparison	Cohen's *d*	95% CI
Experiment 1						
Gain-framing	130	51.61%(33.97%)	88.24%(25.60%)**	$t(128) = 6.98$, $p < .001$	1.23	[0.85, 1.60]
Loss-framing	130	70.06%(35.93%)**	79.34%(35.35%)**	$t(128) = 1.48$, $p = .141$	0.26	[-0.09, 0.61]
Experiment 2						
No corrective feedback	130	61.57%(39.42%)*	86.21%(28.64%)**	$t(128) = 3.99$, $p < .001$	0.71	[0.35, 1.06]
Corrective feedback	130	74.84%(32.62%)**	84.85%(31.78%)**	$t(128) = 1.75$, $p = .083$	0.31	[-0.04, 0.66]

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

* $p < .05$. *$p < .001$.

Figure 1

Average percentage of correct answers in the test phase as a function of the success- and failure-feedback conditions as well as the gain- and loss-framing conditions (Experiment 1, top panel) or the no corrective- and corrective-feedback conditions (Experiment 2, bottom panel). Error bars represent ±1 SE