REVIEWS IN ADVANCE

Learning from Errors

Janet Metcalfe

Department of Psychology, Columbia University, New York, New York 10027; email: jm348@columbia.edu

Annu. Rev. Psychol. 2017. 68:6.1–6.25

The *Annual Review of Psychology* is online at psych.annualreviews.org

This article's doi: 10.1146/annurev-psych-010416-044022

Keywords

errorless learning, generation effect, hypercorrection effect, feedback, after-action review (AAR), error management training (EMT), formative assessment, reconsolidation, prediction error

Abstract

Although error avoidance during learning appears to be the rule in American classrooms, laboratory studies suggest that it may be a counterproductive strategy, at least for neurologically typical students. Experimental investigations indicate that errorful learning followed by corrective feedback is beneficial to learning. Interestingly, the beneficial effects are particularly salient when individuals strongly believe that their error is correct: Errors committed with high confidence are corrected more readily than low-confidence errors. Corrective feedback, including analysis of the reasoning leading up to the mistake, is crucial. Aside from the direct benefit to learners, teachers gain valuable information from errors, and error tolerance encourages students' active, exploratory, generative engagement. If the goal is optimal performance in high-stakes situations, it may be worthwhile to allow and even encourage students to commit and correct errors while they are in low-stakes learning situations rather than to assiduously avoid errors at all costs.

Contents

INTRODUCTION

Nobody wants to make errors in a situation that counts. The consequences of committing such errors can be devastating. If one is performing a piano solo before an audience, controlling a nuclear reactor, taking SATs, making a medical decision, fighting on a battlefield, or giving a lecture, the last thing one wants is an error. This review is not directed at the question of whether errors, in a situation that counts, are good—of course they are not. Rather, the question is how, during initial learning and during practice and preparation for a test that counts, one can best get to a state of performance that is optimal and in which errors will not inadvertently occur just when one needs them least and when they will do the most damage. Should one commit, explore, examine, analyze, and correct errors during learning and practice sessions, or should one avoid errors at all stages of learning?

It might seem intuitive that if one does not want errors on the test that counts, then one should avoid errors at all stages of learning. In this view, committing errors should make those errors more salient and entrench them both into the memory and the operating procedures of the person who makes them. Exercising the errors should make the errors themselves stronger, thus increasing their probability of recurrence. Such a view, which is consistent with a number of the oldest and most well established theories of learning and memory (Bandura 1986, Barnes & Underwood 1959, Skinner 1953), suggests that errors are bad and should be avoided at all costs.

In keeping with these views, Ausubel (1968) warned of the dangers of errors in the learning process and suggested that allowing people to make errors encourages them to practice incorrect and inefficient approaches that will cause trouble because they are difficult to overwrite later with correct approaches. He used this reasoning to argue against an exploratory learning strategy, which by its very nature would mean that incorrect paths and faulty approaches and solutions would be encountered and entertained by the learner. Ausubel (and others) feared that with active exploratory learning, these false starts and errors would be learned and would make learning of the correct solutions and procedures more difficult, if not impossible. Accordingly, active exploratory learning was to be avoided.

Similarly, Bandura (1986, p. 47) urged that learners should be "spared the costs and pain of faulty effort" and that they should instead receive the needed step-by-step guidance that results in flawless behavior from the outset. Feedback should focus only on the correct execution of tasks and should take the form of positive social reinforcement, with errors—if any—being ignored. Indeed, errors can have a detrimental effect for people who have particular neurological deficits; this topic is discussed in further detail below. However, a fear of errors typically diverts learning from highly productive generative strategies, and error avoidance strategy is even more pernicious because generating errors—as long as corrective feedback is given–is actually beneficial to learning.

ENCOURAGING VERSUS DISCOURAGING ERRORS IN THE CLASSROOM

It is not clear how important the worry about avoiding errors has been in shaping American teaching strategies. It is extremely difficult to obtain accurate data concerning what and how teachers teach, let alone to manipulate that teaching in a manner that would allow confident inferences to be made. However, Stevenson & Stigler (1994; see also Stigler & Hiebert 2009) and their colleagues conducted a landmark study in which they were able to videotape lessons in grade 8 mathematics classrooms in a variety of countries, including the United States, Taiwan, China, and Japan. Of most interest, given that Japan is by far outstripping the United States in math scores, is the striking difference in the teaching methods used in those two countries. Although there may be many other reasons for the differences in math scores, one highly salient difference is whether or not teachers engage with students' errors. Videotapes show that, in the United States, set procedures for doing particular kinds of problems are explicitly taught. These correct procedures are rehearsed and emphasized; errors are avoided or ignored. The students are not passive in American classrooms. A teacher may ask for student participation in repeating, for example, a procedure for borrowing when subtracting. When asking a question such as, "Can you subtract 9 from 5?" to prompt students to answer, "No, you have to borrow to make the 5 a 15," the teacher may fail to even acknowledge the deviant child who says, "Yes. It's negative 4." If the response does not fit with the procedure being exercised, it is not reinforced. Errors (as well as deviant correct answers) are neither punished nor discussed but are disregarded. Praise is given, but only for the "correct" answer.

As Stevenson & Stigler (1994) pointed out, praise curtails discussion and serves mainly to reinforce the teacher's role as the authority who bestows rewards. It does not empower students to think, criticize, reconsider, evaluate, and explore their own thought processes. By way of contrast, in Japan praise is rarely given. There, the norm is extended discussion of errors, including the reasons for them and the ways in which they may seem plausible but nevertheless lead to the incorrect answer, as well as discussion of the route and reasons to the correct answer. Such in-depth discussion of the thought processes underlying both actual and potential errors encourages exploratory approaches by students.

Instead of beginning with teacher-directed classwork and explication, Japanese students first try to solve problems on their own, a process that is likely to be filled with false starts. Only after these (usually failed) attempts by students does teacher-directed discussion—interactively involving students and targeting students' initial efforts and core mathematical principles—occur. It is expected that students will struggle and make errors, insofar as they rarely have available a fluent procedure that allows them to solve the problems. Nor are students expected to find the process of learning easy. But the time spent struggling on their own to work out a solution is considered a crucial part of the learning process, as is the discussion with the class when it reconvenes to share the methods, to describe the difficulties and pitfalls as well as the insights, and to provide feedback on the principles at stake as well as the solutions.

As Stevenson & Stigler (1994, p. 193) note, "Perhaps because of the strong influence of behavioristic teaching, which says conditions should be arranged so that the learner avoids errors and makes only a reinforceable response, American teachers place little emphasis on the constructive use of errors as a teaching technique. Learning about what is wrong may hasten understanding of why the correct procedures are appropriate, but errors may also be interpreted as failure. And Americans, reluctant to have such interpretations made of their children's performance, strive to avoid situations where this might happen."

The Japanese active learning approach well reflects the fundamental ideas of a learning-from-errors approach. Engaging with errors is difficult, but difficulty can be desirable for learning (Bjork 2012). In comparison with approaches that stress error avoidance, making training more challenging by allowing false starts and errors followed by feedback, discussion, and correction may ultimately lead to better and more flexible transfer of skills to later critical situations.

ERROR GENERATION AND MEMORY FOR CORRECT RESPONSES IN THE LAB

Considerable research now indicates that engagement with errors fosters the secondary benefits of deep discussion of thought processes and exploratory active learning and that the view that the commission of errors hurts learning of the correct response is incorrect. Indeed, many tightly controlled experimental investigations have now shown that in comparison with error-free study, the generation of errors, as long as it is followed by corrective feedback, results in better memory for the correct response. Oddly, the researchers who have investigated these processes usually refer to errors as "unsuccessful retrieval," leaving ambiguous whether a mistake was produced or whether the learner had just neglected to generate the correct response.

Early studies by Izawa (1967, 1970) showed that multiple unsuccessful retrieval attempts led to better memory for the correct feedback than did a procedure producing fewer incorrect responses.

Kane & Anderson (1978) showed similar results: Attempting the generation of the last word of the sentence, even if what was generated was wrong, led to enhanced correct performance compared to reading the sentence correctly from the outset. Slamecka & Fevreiski (1983) asked people to remember near antonyms, such as trivial-vital or oscillate-settle. Even failed attempts (followed by feedback containing the correct answer) improved later recall of the correct answers over simply reading the correct answer. Kornell et al. (2015) have conducted a recent investigation of the same issue and have reached similar conclusions.

People rarely produced explicit mistakes in the early experiments. The majority of so-called retrieval failures were omission rather than commission errors. However, when Slamecka & Fevreiski (1983) specifically analyzed the effect of commission as compared to omission errors, they found that commission errors, as long as they were semantically related to the target, resulted in better later recall of the correct answer.

Kornell et al. (2009) conducted the first definitive study that directly compared the effect of producing versus not producing a commission error. The cue and the to-be-remembered target word in most of the experiments in their study were slightly related word pairs. They compared a condition in which the answer, or target, was simply given to participants, with no intervening error generation (the no-error condition), to one in which the participants were asked to guess the answer first and nearly always produced an error before being given the correct answer to study (the error-generation condition). The experiment was carefully controlled to ensure that the amount of time spent studying the correct answer was equated across conditions. Kornell et al. (2009) also eliminated from consideration any instances in which the person did not generate an error in the error-generation condition. This happened less than 10% of the time. The surprising finding, which has now been replicated many times, was that on the final test, participants remembered the correct answers considerably better when they had generated an error than when they had not. It appears, then, that error generation is not inevitably bad and to be avoided at all costs. Indeed, error generation appears to foster learning (see **Figure 1**).

A flurry of replications quickly followed. For example, Huelser & Metcalfe (2012) found a beneficial effect of error generation as long as the errors people produced were not completely unrelated to the target. They found no benefit with unrelated pairs—a finding that was replicated by Grimaldi & Karpicke (2012) and by Knight et al. (2012) and that is consistent with Slamecka &

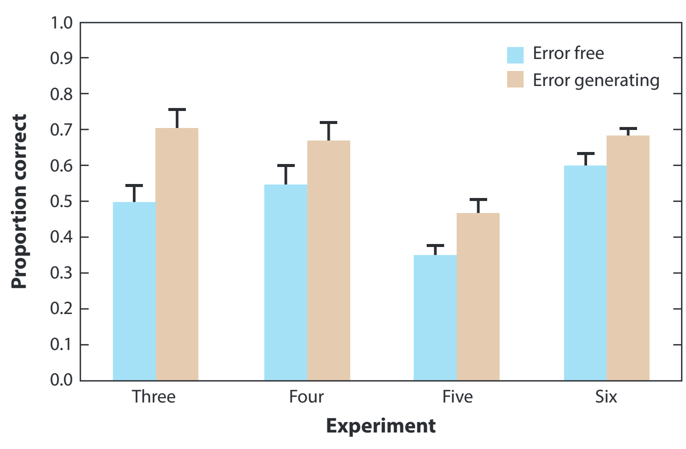

Figure 1

Proportion of correct responses following error-generating and error-free study. Figure adapted with permission from Kornell et al. (2009).

Fevreiski's (1983) earlier analysis. Similarly, Kang et al. (2011) found no effect when participants had absolutely no idea what the answers to factual questions might be but were nevertheless forced to guess. It appears that to be beneficial, the guess needs to be somewhat informed rather than a shot in the dark. Interestingly, in the related-pair case in which a large beneficial effect of committing errors was found, the participants were metacognitively unaware of the benefit. Even immediately after they had experienced the task and had evidenced a benefit of 20% (i.e., roughly the difference between a C− and an A, if it had been a course grade), participants thought that the error-free condition had resulted in better recall (Huelser & Metcalfe 2012). This lack of awareness of the benefits of error generation may contribute to the aversion to errors in the American teaching style evinced in Stigler's work.

In a more educationally realistic study, Richland et al. (2009) found that error generation was related to enhanced memory for material from reading passages. In their study, material from the passage was tested before participants had a chance to read the passage, resulting in participants generating many errors. However, later correct memory for the material that had been pretested in this way was greatly enhanced. More recently, instead of using simple word pairs, Kornell et al. (2015) used general information as their stimuli, asking people either to try to generate the responses (resulting in many mistakes) or to read the question and correct answer. In this situation, too, generation of the error helped later memory for the correct response. Thus, it appears that generating an error, if it is related in any way to the correct response, enhances rather than impairs memory for the correct answer.

Feedback to Errors

When people have made an error, corrective feedback is crucial (Anderson et al. 1971; Butler & Roediger 2008; Hancock et al. 1992; Kornell & Metcalfe 2013; Lhyle & Kulhavy 1987; Metcalfe & Kornell 2007; Metcalfe et al. 2007, 2009; Pashler et al. 2005). It is not enough to simply tell learners whether they were right or wrong. People get virtually no benefit unless the feedback they receive provides the correct answer, as Pashler et al. (2005) showed (and see Bangert-Drowns et al. 1991, Moreno 2004). Furthermore, people need to understand and pay attention to the feedback. Many studies showing a lack of benefit from feedback (and some showing a lack of benefit from committing errors) fail to ensure that the feedback is processed. When the correct answer is made available, though, and people appreciate that the answer is correct as well as why that answer is correct, they are able to integrate that information into memory and improve performance (Anderson et al. 1971).

Furthermore, Metcalfe et al. (2009) showed that the feedback did not have to be given within moments of error commission to be effective. They contrasted a condition in which people were given the feedback immediately upon error commission with one in which it was delayed by up to a week (and they controlled for lag until test, which was one week from feedback in both conditions, so the amount of time until the test did not differ in the two conditions). In both the immediate and delayed feedback conditions, participants were required to attend to the feedback, insofar as they were instructed to type the answer into the computer. The study found that college students performed equally well in the immediate and delayed feedback conditions, whereas children in grades 3 to 5 did better when the feedback was delayed. Interestingly, Kulik & Kulik (1988) noted that whether delayed or immediate feedback produced better results differed between studies conducted in the classroom and those in the laboratory. Lab studies tended to show that delayed feedback was better, whereas classroom studies favored immediate feedback. They concluded, however, that the real difference between these studies was whether the learners paid attention to the feedback. Students in the classroom are highly engaged in knowing the answers to questions right after taking a test. They pay attention to the feedback when it is given immediately. However, their interest flags with a long delay, especially if the feedback is cursory. If the teacher needs or wants to delay feedback (to allow for conscientious scoring, for example), it seems clear that the students' interest in the questions needs to be reignited, and the presentation of the corrective feedback needs to be engaging.

When feedback is elaborative or scaffolded, its beneficial effects are also increased (Finn & Metcalfe 2010). The provision of feedback can be, as it is in the Japanese classrooms, extensive and hence very helpful in allowing students to understand the underlying constructs and to generalize to new situations. Feedback provides an opportunity for exploring and deeply analyzing the principles underlying the problems, the answers, and the reasoning leading to the answers—be they right or wrong. Siegler (1995) conducted a particularly interesting study illustrating the importance of focusing students' attention on the reasoning underlying the feedback. He had three conditions: (*a*) a feedback-only condition; (*b*) an explain-own-reasoning condition, in which the children were asked, "Why do you think that?"; and, most effectively, (*c*) an explain-the-correct reasoning condition, in which the students, after having been given the corrective feedback, were asked, "How do you think I knew that?" As Siegler (2002, p. 40) noted, "Having the children explain another person's correct reasoning has the advantage of both discovery and didactic approaches to instruction. It is like discovery-oriented approaches in that it requires the child to generate a relatively deep analysis of a phenomenon without being told how to do so. It is like didactic approaches in that it focuses the child's attention on the correct reasoning. Thus, it combines some of the efficiency of didactic instruction with some of the motivating properties of discovery."

Sources of Errors

Another fundamental question concerns whether the source of an error modulates its efficacy for learning. The literature indicates that although self-generating an error and being exposed to externally presented irrelevant information may seem similar on the surface, important differences may exist. For example, Grimaldi & Karpicke (2012) showed that error generation, as opposed to merely being presented the correct response (in a paradigm like that of Kornell et al. 2009), had a beneficial effect on later memory for the correct answer. However, when instead of having the participant generate an error, the experimenter presented an incorrect word during what would have been the generation period, memory for the correct answer was harmed. Furthermore, it was also harmful to restrict what the person had to generate as an error (a constrained alternative). A similar phenomenon has been observed with self-generated as compared to experimenter-presented mistaken items when people were in tip-of-the-tongues states. Kornell & Metcalfe (2006) noted that self-generated errors (which, typically, the participants knew were not the correct answers) did not block retrieval of the correct answer, as opposed to when no incorrect answer was produced. In contrast, when experimenter-presented incorrect answers were given while people were in tip-of-the-tongue states, the incorrect answers harmed retrieval of the correct answer in comparison with having nothing presented (Smith & Blankenship 1991). Determining how self-generated wrong answers differ from experimenter-presented wrong items may be important for understanding why error generation helps. It is possible that self-generated errors tap into the person's own semantic memory structure and thus can serve as mediators to help the person get to the correct answer. Externally imposed distracting items may fail to take advantage of an individuals' internal mental structure and thus fail to serve the mediational role. Such items may simply be distracting.

What happens if a person observes another person—say, a fellow student—make an error that is then followed by feedback? Does this exposure have the same beneficial effect as self-generating one's own error, or is it like being exposed to irrelevant experimenter-presented distractions? This is an interesting question with practical implications, but it is an issue that research has not yet addressed.

CONFIDENCE IN ERRORS

Errors Versus Guesses

The critic may argue that perhaps the people in the experiments described above did not really believe in their responses: They were just generating guesses, not genuine errors. Perhaps there is no benefit to generating errors when people really endorse their errors and have strong beliefs that they are correct. The data show greater benefit when the errors were related to the target than when they were unrelated, which somewhat offsets this contention. Even so, few classroom teachers would ask students to memorize lists of word pairs such as Oscillate-Settle or Hillside-Banker or similar items used in the experiments on error generation. Kornell and colleagues' (2015) experiment with real general-information questions—which showed error-generation benefits in a real-world factual task—helps allay some concerns that benefits may obtain only with guesses and not with real errors.

Nevertheless, we may ask how much belief is enough to make a particular response a genuine error rather than just a guess. In the next section, we review the literature on the relation between individuals' degree of belief, or their confidence in the truth of their answers to general-knowledge questions, and their propensity to correct those answers and remember the corrections when they turn out to be erroneous.

Correcting High-Confidence Versus Low-Confidence Errors

A number of studies have investigated the correction of errors as a function of the individual's confidence in the error. Typically, a factual question is asked of the participant; the individual generates an answer and then rates his or her confidence that the error is correct. Next, feedback is given, in which the correct answer is provided. For example, participants might be given the question, "What kind of music is associated with the Cajuns in Louisiana?" Participants then give an answer (perhaps "jazz") and rate their confidence in the correctness of the answer they produced. Finally, participants are given the correct answer (in this case, "zydeco"). In contrast to the predictions of a variety of theories that suggest that responses in which one is highly confident should be particularly difficult to overwrite, the high-confidence errors are more likely to be corrected on the retest than are errors endorsed with lower confidence (e.g., Butler et al. 2008, 2011; Butterfield & Mangels 2003; Butterfield & Metcalfe 2001, 2006; Cyr & Anderson 2013; Eich et al. 2013; Fazio & Marsh 2009, 2010; Iwaki et al. 2013; Kulhavy et al. 1976; Metcalfe et al. 2012; Metcalfe & Finn 2011, 2012; Sitzman & Rhodes 2010; Sitzman et al. 2015). This hypercorrection effect occurs both with immediate retest and when the retest is given at a considerable delay (Butler et al. 2011, Butterfield & Mangels 2003, Metcalfe & Miele 2014). A strong degree of belief in the truth of one's errors makes them more, rather than less, susceptible to being correctable.

Explanations of the Hypercorrection Effect

A number of associative theories of memory and of the relation of memory to confidence seem to indicate that the hypercorrection effect should not be found; indeed, perhaps even the reverse should be observed. Responses that are made with high confidence are those in which the person believes most and are thought to be the strongest in memory (e.g., Ebbesen & Rienick 1998).

As such, they should be most easily accessible and most resistant to interference as well as the most difficult to overwrite or replace with a new response. Certainly, in all data presented to date on the hypercorrection effect, the correlation between confidence in one's first responses and the correctness of those responses is high: The responses in which people are highly confident are nearly always correct, indicating that in general people know what they do and don't know and that their high-confidence responses are strongly and readily retrieved from memory. Most errors that people make are assigned low confidence (and because they are weak, should be easy to overwrite with a new response). People make high-confidence errors only rarely. But if such a high-confidence response were in error, it too—like correct high-confidence responses—should be strong, entrenched, and difficult rather than easy to change. The consistent finding of a hypercorrection effect, whereby high-confidence errors are corrected more readily than low-confidence errors, flies in the face of traditional interference, response hierarchy, and associative theories.

Two nonmutually exclusive factors are central to the hypercorrection phenomenon in young adults, though these are not the only factors at play. The first relates to the surprise individuals experience at being wrong when they were sure they were right. The second relates to the structure of the semantic network surrounding high- as compared to low-confidence errors.

Because they are surprised (and perhaps embarrassed) at having made a mistake on a response they strongly thought was correct, individuals may rally their attentional resources to better remember the correct answer. Several lines of evidence support this surprise/attentional factor. For example, Butterfield & Metcalfe (2006) conducted an experiment designed to evaluate the extent to which one task drew attention from another. Subjects were required to detect when very soft tones occurred while they were doing the general-information error-correction task. Some of the tones were intentionally presented simultaneously with the corrective feedback following high- and low-confidence errors. Participants were more likely to fail to detect the tones that were presented at the time of high-confidence error feedback than those presented at the time of low-confidence error feedback. This result indicates that because participants' attention was captured by the high-confidence error feedback, they had less capacity to detect the tones.

Fazio & Marsh (2009) tested for memory for the surrounding context that accompanied feedback. They showed that the surrounding context of high-confidence error feedback was better remembered than was the surrounding context of low-confidence error feedback, and they interpreted this result as favoring the attentional explanation.

Butterfield & Mangels (2003) used young adult participants in the first event-related brain potential (ERP) study on this paradigm. The most salient result from their study was that participants showed a voltage deflection—the P3a—that past literature on ERPs has linked to surprise reactions in which people rally their attention (Friedman et al. 2001). The literature has also associated this deflection with enhanced memory (Paller et al. 1985). Butterfield & Mangel's (2003) study revealed that there was a confidence-graded P3a when corrective feedback was given, a result that has been replicated and extended by Metcalfe et al. (2015) (see **Figure 2**). The P3a was largest for the feedback to high-confidence errors and smallest for the feedback to low-confidence errors. These results support the contention that high-confidence errors are surprising and that increased attention is paid to corrective feedback to such errors.

Finally, Metcalfe et al. (2012) conducted a study on hypercorrection using event-related functional magnetic resonance imaging. Participants answered questions for several hours, giving their confidence in their answer outside the scanner. Participants then entered the scanner and were presented with questions, their original answers, their original confidence ratings, and finally the correct answer. When the brain activations to the feedback to high- and low-confidence errors were compared, it was found that medial frontal areas that prominently included the anterior cingulate—an area related to surprise, error detection, and attention—were differentially

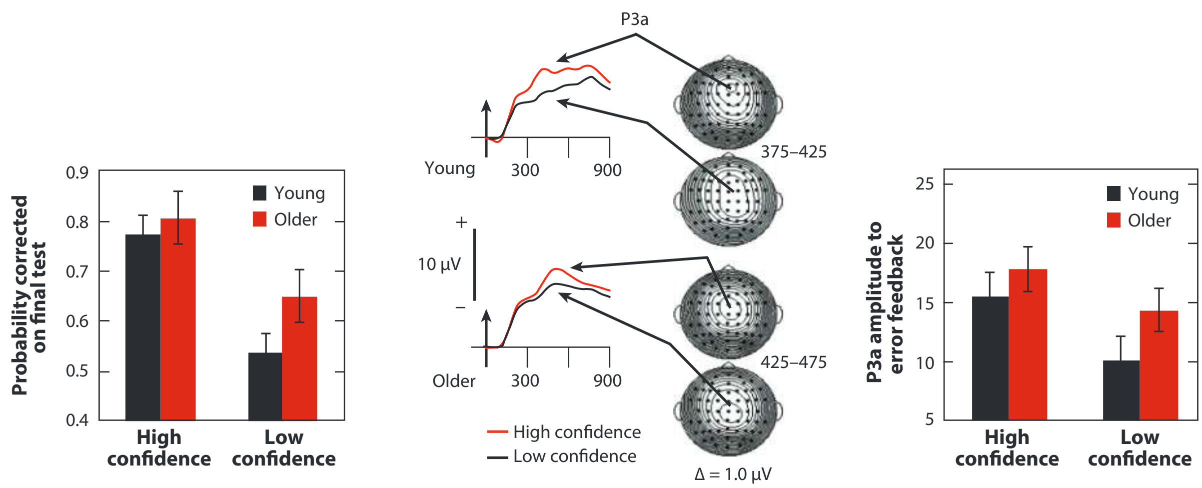

Figure 2

Event-related potentials (ERPs) synchronized to the onset of corrective feedback given to high- and low-confidence errors made by young adults and older adults. The ERP tracings show a prominent P3a to feedback to high-confidence errors. The behavioral results showing the proportion correct on the final test, for both groups as a function of initial error confidence, are shown on the left. Figure adapted with permission from Metcalfe et al. (2015).

activated. Other areas, such as the dorsolateral prefrontal cortex and the temporal parietal junction (see **Figure 3**), were also differentially activated. These results also implicate a surprise-related explanation of the enhanced encoding associated with the feedback to high-confidence errors.

The second factor that has been implicated in the hypercorrection effect is a greater semantic knowledge in the domain of the high-confidence errors than in the low-confidence error domain. Butterfield & Mangels (2003) noted that participant-ascribed familiarity to high- and to

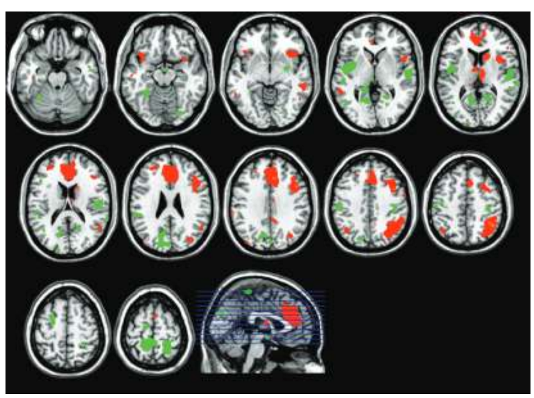

Figure 3

Contrast map of high-confidence errors versus low-confidence errors. Red areas were more active for high-confidence errors, and green areas were more active for low-confidence errors; both areas had an uncorrected threshold of $p < 0.001$. Figure adapted with permission from Metcalfe et al. (2012), © 2012 by the Massachusetts Institute of Technology, **http://www.mitpressjournals.org/doi/abs/10.1162/jocn_a_00228**.

low-confidence responses was differential. Butterfield & Metcalfe (2006) showed that the a priori probability of a correct response was greater for high- than for low-confidence errors.

Latent semantic analysis (Landauer & Dumais 1997) has shown that the similarity of the error to the correct response is higher for high- than for low-confidence errors. This semantic similarity has been found both for young (Metcalfe & Finn 2011) and older (Eich et al. 2013) adults. Thus, individuals are "closer" in semantic space to the right answer when they produce a high-confidence error than when they produce a low-confidence error.

Furthermore, Metcalfe & Finn (2011) have shown that young adults claim that they "knew it all along" frequently when they receive corrective feedback to high-confidence errors. In Metcalfe & Finn's study, young adults were more likely to produce a second guess that was correct following a high- as compared to a low-confidence error. They were more likely to choose the correct alternative in a multiple-choice test that excluded their original answer if the error had been committed with high confidence. Additionally, young adults required fewer clues to guess the correct answers to questions on which they had made high- as compared to low-confidence errors. These results indicate that familiarity with the domain of the high-confidence error, and plausibly with the answer itself, plays a role in the hypercorrection effect.

Children as well as adults showed these familiarity effects (Metcalfe & Finn 2012). Familiarity effects appear to follow the pattern of hypercorrection, but even so, these effects are probably coupled with the effects of surprise. When a person is highly familiar with a domain of knowledge, it is within that domain that surprise at one's errors is most salient.

EXCEPTIONS

Amnesics and Error-Free Learning

People with amnesia present an exception to the general finding that learning from errors is helpful. In a seminal effort to promote learning in people with severe amnesia, Glisky et al. (1986) devised a procedure called the method of vanishing cues (MVC). MVC was developed on the assumption that errors produce enormous interference in amnesics and hence are particularly disruptive for these patients (see, e.g., Hayman et al. 1993). The idea was to always present sufficient information so that the patient never produced a mistake, allowing them to circumvent the proactive interference while at the same time to benefit from priming. Some of the initial demonstrations of MVC were implemented in attempts to teach deeply amnesic patients definitions of computer terms. The experimenters first exposed the definition for 10 seconds, and if the patient failed to produce the word, the first letter appeared on the computer screen. If the patient still failed to produce the target word after 10 more seconds had elapsed, an additional letter was provided, and so on until the word was produced correctly. Then the computer reversed the procedure, providing one less letter on each subsequent trial. Using this method, K.C., a profoundly amnesic patient, was eventually able to "remember" the previously unlearned semantic terms with no letter cues at all. Other researchers (e.g., Baddeley & Wilson 1994, Hamann & Squire 1995; for a review, see Clare & Jones 2008) have also shown that amnesic patients can benefit from errorless learning. These results, while fascinating in their own right, also have implications for theories seeking to explain why making errors helps later correct performance in neurologically typical participants.

Older Adults and Hypercorrection

Older adults provide a different kind of exception. Although older adults benefit from making errors, their pattern of confidence-related benefits is different from that of younger people. Eich and colleagues' (2013) study in which older adults were compared to young adults revealed a typical hypercorrection effect with young adult participants, but the older adults corrected high-confidence errors only slightly more than low-confidence errors (see also Cyr & Anderson 2013). A hypercorrection experiment with older adults by Sitzman et al. (2015) indicated that they had greater familiarity with the low-confidence corrections than did the younger adults—evoking a preexperimental familiarity explanation. Metcalfe et al. (2015) also presented data indicating that older adults claimed greater familiarity with the correct answers to low-confidence errors than did younger adults. Interestingly, though, their results indicate that the older adults may not have actually relied that much on preexperimental familiarity: In comparison with younger adults, they also showed greater learning of answers with which they were unfamiliar. These authors argued for an attentional explanation of the decreased hypercorrection effect because compared to young adults, the older adults showed a larger attention-related event-related P3a to their low-confidence errors. The older adults not only hypercorrected their high-confidence errors but also did extremely well in correcting their low-confidence errors—errors that young adults tended not to correct.

IMPLICATIONS OF THE HYPERCORRECTION EFFECT

Scientific Understanding

Van Loon et al. (2015) investigated the correction of scientific misconceptions, which is one of the most interesting extensions of the hypercorrection effect. They looked at people's correction of conceptual errors, a highly relevant and ecologically valid situation in which error correction is vital for understanding. They found that although hypercorrection did not occur when students simply read standard texts that presented the correct factual information without emphasis on the conceptual errors, it did occur when the students read what they called 'refutation' texts. In these latter texts, the misconceptions that the students had voiced were explicitly stated and then discussed and refuted. The authors suggested that individuals are surprised when they find that a misconception held with high confidence is, in fact, wrong, and the surprise results in enhanced attention to the refutation.

Lexical Representations

Iwaki et al. (2013) extended the hypercorrection paradigm to the correction of erroneous lexical representations found in language learning. In their study, participants performed a Japanese kanji word-reading task, in which the inscriptions had several possible pronunciations, only one of which was correct. Participants read each word aloud and indicated their confidence in their response before being given visual feedback of the correct response. As in the factual knowledge case, a hypercorrection effect was observed in this lexical representation case. Interestingly, the effect was modulated by participants' evaluation of the practical value of to-be-learned words.

Correcting False Inferences in Episodic Memory

Fazio & Marsh (2010) conducted a study in which participants made erroneous inferences, sometimes with high confidence. For example, individuals were asked to remember statements such as "The clumsy chemist had acid on his coat" or "The karate champion hit the cinder block." Later, the participants often made inferences beyond what was actually given in the statement. For example, they might have recalled, "The clumsy chemist *spilled* acid on his coat," or "The karate champion *broke* the cinder block." The authors asked participants to rate confidence in their answers and then gave feedback concerning the original statement. Interestingly, a hypercorrection effect was found in this ecologically valid situation—the greater participants' confidence was in their wrong answers, the greater the correction upon retest. This study suggests that false episodic memories, in which people often express high confidence, may be open to rectification when the correct information is provided.

THEORIES OF WHY ERRORS ENHANCE LEARNING

As was noted in the introduction, standard interference theory does a poor job of accounting for why generating errors helps later memory for the correct answer. Other theories, though, do better.

Mediation

In seeking to explain why testing helps learning, Pyc & Rawson (2010) and Carpenter (2011) proposed that when individuals are tested, they are likely to generate effective mediational retrieval cues. An interesting possibility is that erroneous responses serve as signposts or stepping-stones rather than as interfering competitors to the correct answer. This conjecture is consistent with the data showing that only errors that are related to the target are effective in aiding later recall. It is also consistent with the idea that the errors have to be self-generated to help. Externally presented items seem, intuitively, much less likely to tap into the individual's own semantic memory structure in such a way that they could later provide a route to the target correct item. Whether such mediation might be entirely semantic or might rely on an episodic-memory component has not been fully explored. However, a position outlined by Jacoby & Wahlheim (2013) suggests the latter.

Recursive Reminding

Jacoby & Wahlheim (2013) discussed the possibility that errors (and other contextual aspects of the encoding situation) have a facilitative effect to the extent that they are related to retrieving the original episodic event in which both the correct answer and the mistake were embedded. The idea is that individuals may be able to remember the context in which they made an error and by so doing think not only of the error itself but also of the surrounding context in which it was made clear that the error had been made and that the correction was the desired answer. This process, which is called recursive reminding, relies on the learner's explicit or episodic memory for the event, including making the mistake and receiving the correction. This explanation of the beneficial effects of errors is particularly attractive in light of the findings that people with amnesia, who have impaired explicit or episodic memory, are harmed by generating errors, whereas people with intact episodic memory benefit from making errors. However, not all data are supportive. The theory would seem to imply that memory for the correct answer should be particularly enhanced if the error is regenerated at time of retrieval. However, Butterfield & Metcalfe (2001) and Metcalfe & Miele (2014) showed that correct recall was independent of error generation at time of retrieval. On the other hand, Wahlheim & Jacoby (2013) showed a positive relation between remembering that there had been a change from what the person had previously thought was the answer, at time of encoding, and later target memory. Clearly, further research is needed to test this promising theory.

Reconsolidation

The reconsolidation framework was formulated within a fear-conditioning paradigm that is entirely different from the present error-correction paradigm (e.g., Dudai 2012, Lee 2008, Nader et al. 2000). The reconsolidation framework has been directed at discovering methods to overcome posttraumatic stress disorder and traumatic conditioned fears (see Schiller et al. 2010) and is compatible with a number of therapies. Despite a difference in domain, striking similarities exist between the reconsolidation framework and the error-correction paradigm.

The basic premise is that in order for a conditioned fear to be altered or eradicated, the dysfunctional response first needs to be evoked. Following fear response retrieval there is a short time window during which the undesirable response can be eradicated or modified and reconsolidated. If the response is not evoked, though, then it stays buried and unchanged. Some similarities between fear conditioning and error correction suggest that applying reconsolidation theory to both domains may be fruitful. The most important of these similarities is the core idea that the undesired response, be it the dysfunctional fear response or the semantic error, needs to be retrieved in order for it to be rendered susceptible to change.

The reconsolidation framework is consistent with the gist of the results of the many studies that have shown that the probability of producing the correct answer is greater when the error was first retrieved in conjunction with the new, overwriting stimulus, as contrasted to when the preexisting error was not evoked and only the correct answer was provided. Interestingly, reconsolidation theory has detailed a number of factors of interest that have not been explored within the semantic error-correction paradigm. First, the temporal window of opportunity for modification of the erroneous response has not been much explored in the semantic memory situation. If such a window exists, then it would be important—for educational reasons—to identify it. Second, in the reconsolidation framework, the more strongly the fear is evoked, the more likely it is to be malleable (Lee 2008). This finding may be analogous to some of the similarity relations seen in the error-correction paradigm and perhaps even to the basic hypercorrection effect, though these relations—and the reasons for them—need to be specified in more detail. Third, in reconsolidation theory, some experiments use neurological agents such as anisomycin, or even electro-shock, to eliminate the fear, which raises the issue of the need for unlearning. Does the semantic error need to be obliterated for a new correct response to be learned? Or, as in some theories such as the recursive reminding model and the mediational model (described above), is the continued mental presence of the error actually helpful?

Finally, is the ouster of the dysfunctional response permanent? Within the fear conditioning/animal-learning domain, there is a long history of empirical investigations of extinction methods that later result in spontaneous recovery. With reconsolidation, however, the learning of the new correct response is thought to be permanent. Only a few studies in the error-correction paradigm have investigated the important analogous question of the potential return of the errors (Butler et al. 2011, Metcalfe & Miele 2014), and the results are not yet definitive. The permanence of error correction is clearly a crucial question for the efficacy of this method in education. Further effort is needed to investigate whether the kind of permanent change thought to obtain with reconsolidation is also found in the domain of error correction.

Prediction Error

The notion that the discrepancy between a person's expectation and the actual outcome is crucial for learning—at all levels from perception to cognition and memory—has been postulated in many neurally based computational and machine models of learning (Friston 2005, Rumelhart &

McClelland 1986). In these models, if everything happens just as is expected, no learning need take place. It is only when expectations are violated or when there is prediction error that the network needs to change to accommodate the unexpected results. The greater the prediction error, the greater the learning. This attractive and parsimonious theory is consistent with many of the findings cited above.

Furthermore, prediction error models seem at first blush to be particularly applicable to the hypercorrection situation insofar as the magnitude of the prediction error determines the extent of learning. Intuitively, high-confidence errors seem like big, serious errors. As is evidenced by the ERP and the attentional literatures described above, high-confidence errors preferentially evoke such a surprise/attentional reaction. Thus, this aspect of prediction error models fits nicely.

A problem needs to be solved, however. If predication error is assessed in terms of the representational characteristics of the item that is retrieved as compared to the correct answer—as seems plausible—then it would appear that the magnitude of the prediction error in the high-confidence error case is smaller than it is in the low-confidence error case, and therefore less learning should occur. Latent semantic analysis (Landauer & Dumais 1997) showed that the similarity between the correct answers and high-confidence errors was greater than the similarity between the correct answers and the low-confidence errors (Metcalfe & Finn 2011; for a model of the experimental results based on high similarity rather than high prediction error, see Sitzman et al. 2015). Furthermore, as noted above, when individuals were simply told they were wrong and were asked to make a second guess, they were more likely to come up with the correct answer when they had made a high- rather than a low-confidence error; they were more able to choose the correct answer in a multiple choice test; and they were more likely to complete the answer correctly when given fragmentary cues (Metcalfe & Finn 2011). Prediction error models have tremendous appeal, and they fit with the apparent surprise that people feel when they are highly confident but wrong. However, if the prediction error framework is to seamlessly apply, theoretic resolution is needed for the mismatch between individuals' apparent surprise at being wrong in high-confidence errors and their equally evident nearness to being right. A formal implementation of the prediction error framework to this particular error correction situation—a situation of enormous practical import—could greatly contribute to our understanding of error correction in the real world and to the predication error framework.

SECONDARY BENEFITS OF ENCOURAGING ERRORS

Active Generation Is Beneficial to Memory for Correct Responses

One of the spinoffs of pedagogical techniques that allow or even encourage students to generate errors is that during such active participation students frequently generate the correct responses. Active generation as opposed to passive study (Slamecka & Graf 1978; for a review, see Bertsch et al. 2007) and the related testing effect (e.g., McDaniel et al. 2007; Roediger & Karpicke 2006a,b) both produce large beneficial effects on learning. Allowing students to self-generate induces one of the strongest beneficial effects on learning in the cognitive literature.

Knowledge of Student Errors Helps the Teacher

The nature of the errors that students generate, when they are allowed to do so, can provide crucial information to the teacher. Much research on formative assessment (Black & Wiliam 1998; see also Dunn & Mulvenon 2009) indicates that any assessment that could be potentially used by teachers to focus instruction is of great interest. This perspective, although not specifically dwelling on errors, is highly compatible with the approach taken here: Errors indicate areas of difficulty for students and as such are the very areas on which the instruction should be focused. Errors also provide indications to the teacher of what students are thinking and what diverts them from the correct solution. Understanding students' misunderstanding can help the teacher focus on the aspects of the to-be-learned concepts that need to be clarified.

Consideration of Errors Offsets Overconfidence

Encouraging students to consider ways in which they could be wrong has a metacognitive benefit of helping to train subjective confidence (see Bjork 1994). The many experimental studies investigating the confidence of individuals reveal an overarching finding of overconfidence (for a review, see Dunlosky & Metcalfe 2009): People believe that they have learned and know things when they do not yet know them (Metcalfe 1998). Furthermore, people's confidence affects their study choices and time (Finn 2008, Metcalfe & Finn 2008). When people have high confidence that they have learned materials, they decline further study. Similarly, when they have high confidence on tests, they may leave a question prematurely, possibly committing a correctible mistake. Pervasive overconfidence, then, has a deleterious effect on both study and eventual test performance.

One method that has been found to be effective in offsetting overconfidence is to have people consider possible errors and think about how they could be wrong (Koriat et al. 1980). Note that the tendency of individuals to assume that anything retrieved is correct (and to disregard the possibility of error) fits with the mechanisms used for making metacognitive judgments (see, for example, Koriat 2008, 2012) and underlies the need for training subjective as well as objective experience.

A reflective practice method used in medical schools to improve diagnostic decisions (Mamede & Schmidt 2004, Schiff et al. 2009) has the students imagine that they have already made their diagnosis on a particular case and that treatment and the course of action, based on that diagnosis, have already taken place. Students then learn that the patient has died. Students are asked to reevaluate their diagnosis on the basis of this outcome and, if warranted, to make a different diagnosis. This method of checking is termed a premortem analysis. A similar method referred to as prospective hindsight (Mitchell et al. 1989) uses consideration of a possible erroneous outcome to improve choice selection. It, too, has been shown to have positive effects on performance.

ORIGIN OF THE IDEA THAT ERRORLESS LEARNING IS A GOOD THING

With all of this evidence that generating errors helps learning, one might wonder where we got the idea—so entrenched in our educational system—that error generation is bad? The classic paper on errorless learning is that of Terrace (1963). He showed that pigeons could be taught to discriminatively peck a red circle as opposed to a green circle by being reinforced in such a way that they never pecked the green circle at all, that is, the pigeons performed in an errorless manner. This method involved first contrasting the reinforced red circle with a neutral grey background (rather than with the green circle); the neutral grey background would be flashed very quickly so that the pigeon would have no time to peck it even if it was so inclined. But the on-time was gradually increased, the neutral grey background color was slowly replaced with a very faint green circle that over trials gradually became more intensely green, and finally both the red and the green circle were fully visible for an amount of time sufficient for the pigeon to peck either stimulus. With this method, the pigeon pecked only the red circle, never making the error of pecking the green circle. The contrasting method involved presenting both the red and green circles at full intensity, with sufficient pecking time from the outset but reinforcing only pecking of the red circle, a procedure that evinced many pecks to the green circle (errors) during training.

In developing the errorless method of learning, Terrace was endeavoring to test a theory of the role of inhibition in learning and was not attempting to develop—nor did he claim to have developed—a better way to learn. However, the method was vaunted by others and put to use in Skinner's teaching machines. The consequences of the two procedures differed, though. For example, some complex differences in later extinction depended on whether the bird learned in an errorless or errorful way and on which initial reinforcement schedule was used. It is not clear how these differences would apply to classroom learning. Other differences between errorful and errorless learning were observed, but it was not clear that errorless learning was better, even in this very narrow paradigm.

First, the pigeon's reaction time to the red circles was faster when the learning procedure had been errorful rather than error free (Terrace 1968). This finding is consistent with the idea that learning was stronger, even in this paradigm, in the errorful condition. Second, the pigeons did not show a directional shift in preferences away from the error after having been trained in an errorless manner. Avoidance or inhibition of the error might be considered a desirable repercussion of learning. This favorable concomitant occurred only in the errorful learning condition (Terrace 1966, 1968). Third, the pigeons did not show a contrast effect, whereby the presence of the green stimulus enhanced responding to the red, in the errorless learning condition, though they did show a contrast effect in the errorful learning condition (Terrace 2010). Enhanced responding to the correct answer is desirable. The actual results stemming from the method of errorless learning, then, appear to contrast with the ideological beliefs concerning the desirability of errorless learning, even within the original context of simplified reinforcement learning.

Were there any consequences that may have been unfavorable for errorful learning? The most salient, as emphasized by Skinner, was that the pigeons did not show gross emotional responses to the green circle (S−) when they were taught in an errorless manner: "The pigeon quietly waits for the next appearance of S+" (Terrace 1966, p. 316). By way of contrast, following discrimination learning with errors, the green circle evoked "various emotional responses such as wing flapping and turning away from the stimulus" (p. 316).

EMOTIONAL CONSEQUENCES OF ERRORS

One possible negative consequence of learning with errors is exaggerated emotionality. This section provides an overview of the literature on this issue.

Event-Related Potential Studies

Many studies have examined an ERP termed error-related negativity (ERN), which is produced when individuals make, and recognize that they have made, a simple motor mistake in a paradigm in which participants are told to press a button whenever an X, say, is presented on-screen and to withhold responding if anything else—say a V—is presented. Every now and then participants wrongly hit the button when the V is presented. When they do, an early negative voltage deflection is observed in the event-related tracing (Gehring et al. 1990) that has been thought to be a marker of unpleasant emotion. If individuals are shown an unpleasant picture just before making such an error, the ERN amplitude is increased (Wiswede et al. 2009). If they are anxious, the ERN amplitude is increased (Proudfit et al. 2013). Furthermore, adolescents diagnosed with anxiety disorder show larger ERNs (Ladouceur et al. 2006).

The effects, though, are complex and depend on task expectations. In addition, it is not clear whether the detection of these simple incorrect motor responses are the same as, or even related to, the responses that people have to cognitive errors. Butterfield & Mangels (2003) did report a fronto-central negativity similar to the ERN in the hypercorrection paradigm. This negativity, though, was not correlated in any discernible way with error correction.

A second ERP voltage deflection—the P3a—has also been associated with the correction of high-confidence errors (Butterfield & Mangels 2003, Metcalfe et al. 2015). The rich literature on this ERP deflection shows its relation to novelty, memory, and attention (e.g., Friedman et al. 2001, Paller et al. 1985), and it may have a relation to emotion, insofar as participants' anxiety can sometimes affect the amplitude of this component (Luck & Kappenman 2011). Consistent with the localization results of the ERP studies on the P3a in error correction, a functional magnetic resonance imaging study (Metcalfe et al. 2012) showed anterior cingulate and mid-frontal activation in conjunction with the presentation of corrections to high- as compared to low-confidence errors. However, peak activation was in areas most strongly associated with cognition rather than emotion (as shown in **Figure 3**) (see Bush et al. 2000).

Emotional Consequences of Medical and Police Errors

When the consequences of errors are severe, such as in a misdiagnosis or other medical error, the emotional consequences for the agent can also be severe (Wu 2000). Many published studies (e.g., Delbanco & Bell 2007, Waterman et al. 2007) document extreme and troubling emotional consequences felt by doctors. Similarly, police errors can result in severe emotional consequences (Blum & Poliscar 2004). These are unequivocally mistakes on the test that counts and have terrible consequences for another human being for whom the doctor or the police officer feels responsible. As such, these errors can result in guilt feelings and considerable emotion. However, they are not the kind of errors that occur in an inconsequential practice situation, which is the type we focus on in this review.

Rarely have the emotional consequences of making conceptual or factual mistakes in the field (but not in a life-threatening situation) been studied experimentally, although such consequences have been assumed to exist. One exception was an applied managerial study by Zhao (2011), in which it was found that the effect associated with errors was helpful rather than harmful to learning and motivation.

Personality Differences in Responsivity to Errors

Higgins (1999) has proposed that the regulatory focus of individuals—whether they are primarily prevention focused or promotion focused—is related to their sensitivity to errors (Scholar et al. 2008). People who are promotion focused are concerned with advancement, growth, and achievement and care little about making errors; people who are prevention focused are concerned with safety, security, and responsibility and are concerned not to make mistakes. Within a signal-detection framework, the promotion-focused person is concerned with maximizing hits and to do so is willing to tolerate false alarms (i.e., errors). A prevention-focused person is unwilling to tolerate mistakes, even if it means missing out on opportunities. Although regulatory focus is often seen as a dispositional tendency, it can also be experimentally manipulated. Crowe & Higgins (1997) found that failure on a difficult task was more demotivating for people in the prevention set: They produced fewer solutions in a problem-solving task presumably because they monitored their output more closely, and they were wary of producing errors. In contrast, promotion-focused individuals showed a risky bias. It is not known, however, whether prevention-focused or promotion-focused individuals learn better. Metcalfe & Miele (2014) showed subtle effects of being prevention focused versus promotion focused in the hypercorrection paradigm. Prevention-focused individuals showed a larger hypercorrection effect when they were tested immediately, but the effect did not persist at a delay, and overall, there were no learning differences.

Differences in sensitivity to errors among individuals are proposed in other theories of personality as well. Dweck (2006), in her "implicit theories of intelligence" framework, proposed that people may be classified as being either incremental theorists or entity theorists, depending upon their stance toward the malleability of intelligence. Incremental theorists are growth oriented and believe that intelligence is malleable. They have a distinctly positive attitude toward errors, which, for them, provide an opportunity for learning. This attitude it thought to render them resilient to setbacks and to have beneficial effects on learning. In contrast, entity theorists believe that intelligence is fixed. Errors, for an entity theorist, are a mark of deficient intelligence and to be avoided at all costs. The entity theorist's stance seems consistent with Skinner's original concerns.

Methods to Buffer Potential Negative Emotional Effects

Although a number of studies are relevant to overcoming potentially negative emotional effects of errors on motivation, perhaps the most significant are those that have investigated error management training (EMT), an implementation used in software training (see Keith & Frese 2008). EMT participants are explicitly encouraged to make errors during training and to learn from them. They are also given error management instructions, which tell them to expect errors while they work and emphasize the positive informational feedback of errors for learning. EMT has been contrasted with proceduralized training methods, which mimic conventional training tutorials that have a negative attitude toward errors. In the proceduralized condition, detailed step-by-step instructions on correct task solutions are provided to prevent participants from making errors. To investigate the conditions under which EMT does and does not provide learning benefits, Keith & Frese (2008) conducted a meta-analysis of 24 EMT studies. Most of the studies were in the area of software skills, although three investigated decision-making tasks. The results were highly favorable to EMT, with the beneficial effects being particularly favorable for posttraining transfer. This benefit is important because transfer to a nonidentical situation, rather than just easy fluency during the lesson itself, is essential for education. The meta-analysis also showed that the benefits of EMT were modulated by an attitude that errors can be good and provide an opportunity for learning. Heimbeck et al. (2003) explicitly tested the role of error attitude instructions and found that the beneficial effect of EMT was improved when participants were given instructions designed to counteract the negative emotional connotations of errors.

Similarly, Dweck & Leggett (1988) have shown that an incremental (error-resilient) mindset can be induced, even among people who might otherwise adopt an entity theory and be error averse. Dweck and colleagues (e.g., Blackwell et al. 2007) have shown that learning is thereby enhanced. Finally, work on stereotype threat suggests that minority students may be particularly vulnerable to the emotional consequences of errors. Cohen et al. (2009), though, have created a self-affirmation manipulation that has been shown to reduce such stresses and enhance performance. Thus, although some individuals may experience emotional reactions from making errors, especially if the teaching method is not sensitive, it appears to be possible to offset these potentially negative consequences.

USING ERRORS TO IMPROVE LEARNING

Although few methods that capitalize on people's errors have been implemented in teaching, one interesting example exists: the American armed forces use a battle training method that focuses on errors. After-action review (AAR; sometimes termed hotwashing) is an army training method described by Druckman & Bjork (1994) in a National Academy of Sciences report (see also Morrison & Meliza 1999). The gist is that units of novice trainees engage in an action against a so-called opposing force that is extremely well trained and well versed. The novice unit typically is defeated decisively. The novices then undergo the AAR, in which rank is put aside and everyone is free to voice their view. A trained facilitator is present to provide helpful specialized knowledge and to ensure that everyone participates and that the sessions are nonjudgmental of the people involved. Every misstep is analyzed, and discussion centers around what people should have done and why. The concept is that more is learned from defeat than from victory, as long as the trainees have a chance to consider all of their mistakes and make plans to remedy them, and that such learning is better accomplished in a simulated battle rather than in actual battle. The next time the unit meets the opposing force, performance improves, and this encounter is also followed by an AAR. Only after many such battles, each followed by an AAR, is the unit ready for deployment. Within the military, there is much praise for this method. General Gordon R. Sullivan asserted that the "AAR as an essential part of training is one of the most important training interventions ever" (quoted in Morrison & Meliza 1999, p. 21). "Managed and conducted by those closest to the activity, AARs identify how to correct deficiencies, sustain strengths, and focus on improved performance of specific tasks, activities, events, or programs" (USAID 2006, p. iii). Although there is little experimental research on this method, there appears, nevertheless, to be considerable consensus about its efficacy. Interestingly, the manual on AAR (USAID 2006) cites extensive evidence from the literature on human memory and cognition that provides supportive data for each aspect of the method.

CONCLUSION

An unwarranted reluctance to engage with errors may have held back American education. The behavioral and neurological data reviewed here indicate that, as long as one is not amnesic, making errors can greatly facilitate new learning. Errors enhance later memory for and generation of the correct responses, facilitate active learning, stimulate the learner to direct attention appropriately, and inform the teacher of where to focus teaching. The concern that errors might evoke dysfunctional emotional reactions appears to be exaggerated. Of course, sensitive handling of errors and avoiding gratuitous punishments—verbal or otherwise—is essential. The research reviewed here suggests that teachers and learners alike should be encouraged to be open to mistakes and to actively use them in becoming prepared for the test that counts.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I am grateful to Robert A. Bjork and Nate Kornell for inspiration. I thank the Institute of Education Science of the US Department of Education for facilitating this work under grant number R305A150467. The views expressed in this review are those of the author.

LITERATURE CITED

Anderson RC, Kulhavy RM, Andre T. 1971. Conditions under which feedback facilitates learning from programmed lessons. *J. Educ. Psychol.* 63:186–88

Ausubel DP. 1968. *Educational Psychology: A Cognitive View*. New York: Holt, Rinehart & Winston

Baddeley A, Wilson BA. 1994. When implicit learning fails: amnesia and the problem of error elimination. *Neuropsychologia* 32:53–68

Bandura A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall

Bangert-Drowns RL, Kulik C-LC, Kulik JA, Morgan M. 1991. The instructional effect of feedback in test-like events. *Rev. Educ. Res.* 61:213–38

Barnes JM, Underwood BJ. 1959. Fate of first-list associations in transfer theory. *J. Exp. Psychol.* 58:97–105

Bertsch S, Pesta BJ, Wiscott R, McDaniel MA. 2007. The generation effect: a meta-analytic review. *Mem. Cogn.* 35:201–10

Bjork RA. 1994. Memory and metamemory considerations in the training of human beings. In *Metacognition: Knowing About Knowing*, ed. J Metcalfe, A Shimamura, pp. 185–205. Cambridge, MA: MIT Press

Bjork RA. 2012. Desirable difficulties perspective on learning. In *Encyclopedia of the Mind*, ed. H Pashler, pp. 242–44. Thousand Oaks, CA: Sage

Black P, Wiliam D. 1998. Assessment and classroom learning. *Assess. Educ.* 5:7–74

Blackwell L, Trzesniewski K, Dweck CS. 2007. Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. *Child Dev*. 78:246–63

Blum LM, Poliscar JM. 2004. Why things go wrong in police work. *Police Chief* 71:49–52

Bush G, Luu P, Posner MI. 2000. Cognitive and emotional influences in anterior cingulate. *Trends Cogn. Sci.* 4:215–22

Butler AC, Fazio LF, Marsh EJ. 2011. The hypercorrection effect persists over a week, but high confidence errors return. *Psychon. Bull. Rev.* 18:1238–44

Butler AC, Karpicke JD, Roediger HL. 2008. Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *J. Exp. Psychol.: Learn. Mem. Cogn.* 34:918–28

Butler AC, Roediger HL. 2008. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem. Cogn.* 36:604–16

Butterfield B, Mangels JA. 2003. Neural correlates of error detection and correction in a semantic retrieval task. *Cogn. Brain Res.* 17:793–817

Butterfield B, Metcalfe J. 2001. Errors committed with high confidence are hypercorrected. *J. Exp. Psychol.: Learn. Mem. Cogn.* 27:1491–94

Butterfield B, Metcalfe J. 2006. The correction of errors committed with high confidence. *Metacogn. Learn.* 1:69–84

Carpenter SK. 2011. Semantic information activated during retrieval contributes to later retention: support for the mediator effectiveness hypothesis of the testing effect. *J. Exp. Psychol.: Learn. Mem. Cogn.* 37:1547–52

Clare L, Jones RSP. 2008. Errorless learning in the rehabilitation of memory impairment: a critical review. *Neuropsychol. Rev.* 18:1–23

Cohen GL, Garcia J, Purdie-Vaughns V, Apfel N, Brzustoski P. 2009. Recursive processes in self-affirmation: intervening to close the minority achievement gap. *Science* 324:400–3

Crowe E, Higgins ET. 1997. Regulatory focus and strategic inclinations: promotion and prevention in decision-making. *Org. Behav. Hum. Decis. Proc.* 69:117–32

Cyr A-A, Anderson ND. 2013. Updating misconceptions: effects of age and confidence. *Psychon. Bull. Rev.* 20:574–80

Delbanco T, Bell SK. 2007. Guilty, afraid, and alone—struggling with medical error. *N. Engl. J. Med.* 357:1682–83

Druckman D, Bjork RA. 1994. *Learning, Remembering, Believing: Enhancing Human Performance*. Washington, DC: Natl. Acad. Press

Dudai Y. 2012. The restless engram: consolidations never end. *Annu. Rev. Neurosci.* 35:227–47

Dunlosky J, Metcalfe J. 2009. *Metacognition*. Thousand Oaks, CA: Sage

Dunn KE, Mulvenon SW. 2009. A critical review of research on formative assessment: the limited scientific evidence of the impact of formative assessment in education. *Pract. Assess. Res. Eval.* 14:1–11

Dweck CS. 2006. *Mindset*. New York: Random House

Dweck CS, Leggett EL. 1988. A social-cognitive approach to motivation and personality *Psychol. Rev.* 95:256–73

Ebbesen EB, Rienick CB. 1998. Retention interval and eyewitness memory for events and personally identifying attributes. *J. Appl. Psychol.* 83:745–62

Eich TS, Stern Y, Metcalfe J. 2013. The hypercorrection effect in younger and older adults. *Aging Neuropsychol. Cogn.* 20:511–21

Fazio LK, Marsh EJ. 2009. Surprising feedback improves later memory. *Psychon. Bull. Rev.* 16:88–92

Fazio LK, Marsh EJ. 2010. Correcting false memories. *Psychol. Sci.* 21:801–3

Finn B. 2008. Framing effects on metacognitive monitoring and control. *Mem. Cogn.* 36:813–21

Finn B, Metcalfe J. 2010. Scaffolding feedback to maximize long-term error correction. *Mem. Cogn.* 38:951–61

Friedman D, Cycowicz Y, Gaeta H. 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci. Biobehav. Rev.* 25:355–73

Friston K. 2005. A theory of cortical responses. *Philos. Trans. R. Soc. B* 360:815–83

Gehring WJ, Coles MGH, Meyer DE, Donchin E. 1990. The error-related negativity: an event-related brain potential accompanying errors. *Psychophysiology* 27:S34

Glisky EL, Schacter DL, Tulving E. 1986. Learning and retention of computer related vocabulary in memory-impaired patients: method of vanishing cues. *J. Clin. Exp. Neuropsychol.* 8:292–312

Grimaldi PJ, Karpicke JD. 2012. When and why do retrieval attempts enhance subsequent encoding? *Mem. Cogn.* 40:505–13

Hamann SB, Squire LR. 1995. On the acquisition of new declarative knowledge in amnesia. *Behav. Neurosci.* 109:1027–44

Hancock TE, Stock WA, Kulhavy RW. 1992. Predicting feedback effects from response-certitude estimates. *Bull. Psychon. Soc.* 30:173–76

Hayman CA, MacDonald CA, Tulving E. 1993. The role of repetition and associative interference in new semantic learning in amnesia: a case experiment. *J. Cogn. Neurosci.* 5:375–89

Heimbeck D, Frese M, Sonnentag S, Keith N. 2003. Integrating errors into the training process: the function of error management instructions and the role of goal orientation. *Personal. Psychol.* 56:333–61

Higgins ET. 1999. Promotion and prevention as a motivational duality: implications for evaluative processes. In *Dual-Process Theories in Social Psychology*, ed. S Chaiken, Y Trope, pp. 503–25. New York: Guilford

Huelser BJ, Metcalfe J. 2012. Masking related errors facilitates learning, but learners do not know it. *Mem. Cogn.* 40:514–27

Iwaki N, Matsushima H, Kodaira K. 2013. Hypercorrection of high confidence errors in lexical representations. *Percept. Motor Skills* 117:219–35

Izawa C. 1967. Function of test trials in paired-associate learning. *J. Exp. Psychol.* 76:194–209

Izawa C. 1970. Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *J. Exp. Psychol.* 83:340–44

Jacoby LL, Wahlheim CN. 2013. On the importance of looking back: the role of recursive remindings in recency judgments and cued recall. *Mem. Cogn.* 41:625–37

Kane JH, Anderson RC. 1978. Depth of processing and interference effects in the learning and remembering of sentences. *J. Educ. Psychol.* 70:626–35

Kang SHK, Pashler H, Cepeda NJ, Rohrer D, Carpenter SK, Mozer MC. 2011. Does incorrect guessing impair fact learning? *J. Educ. Psychol.* 131:48–59

Keith N, Frese M. 2008. Effectiveness of error management training: a meta-analysis. *J. Appl. Psychol.* 93:59–69

Knight JB, Ball BH, Brewer GA, DeWitt MR, Marsh RL. 2012. Testing unsuccessfully: a specification of the underlying mechanisms supporting its influence on retention. *J. Mem. Lang.* 66:731–46

Koriat A. 2008. Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Mem. Cogn.* 36:416–28

Koriat A. 2012. The self-consistency model of subjective confidence. *Psychol. Rev.* 119:80–113
Koriat A, Lichtenstein S, Fischhoff B. 1980. Reasons for confidence. *J. Exp. Psychol.: Hum. Learn. Mem.* 6:107–18
Kornell N, Hays MJ, Bjork RA. 2009. Unsuccessful retrieval attempts enhance subsequent learning. *J. Exp. Psychol.: Learn. Mem. Cogn.* 35:989–98
Kornell N, Klein PJ, Rawson KA. 2015. Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. *J. Exp. Psychol.: Learn. Mem. Cogn.* 41:283–94
Kornell N, Metcalfe J. 2006. Blockers do not block recall in tip-of-the-tongue states. *Metacogn. Learn.* 1:248–61
Kornell N, Metcalfe J. 2013. The effects of memory retrieval, errors, and feedback on learning. In *Applying Science of Learning in Education: Infusing Psychological Science into the Curriculum*, ed. VA Benassi, CE Overson, CM Hakala. Washington, DC: Am. Psychol. Assoc. Soc. Teach. Psychol. **http://teachpsych.org/ebooks/asle2014/index.php**
Kulhavy RW, Yekovich FR, Dyer JW. 1976. Feedback and response confidence. *J. Educ. Psychol.* 68:522–28
Kulik JA, Kulik C-LC. 1988. Timing of feedback and verbal learning. *Rev. Educ. Res.* 58:79–97
Ladouceur CD, Dahl RE, Birmaher B, Axelson DA, Ryan ND. 2006. Increased error-related negativity (ERN) in childhood anxiety disorder. *J. Child Psychol. Psychiatry* 47:1073–82
Landauer TK, Dumais ST. 1997. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104:211–40
Lee JLC. 2008. Memory reconsolidation mediates the strengthening of memories by additional learning. *Nat. Neurosci.* 11:1264–66
Lhyle KG, Kulhavy RW. 1987. Feedback processing and error correction. *J. Educ. Psychol.* 79:320–22
Luck SJ, Kappenman ES. 2011. *The Oxford Handbook of Event-Related Potential Components*. New York: Oxford Univ. Press
Mamede S, Schmidt HG. 2004. The structure of reflective practice in medicine. *Med. Educ.* 38:1302–8
McDaniel MA, Roediger HL, McDermott KB. 2007. Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon. Bull. Rev.* 14:200–6
Metcalfe J. 1998. Cognitive optimism: self-deception of memory-based processing heuristics. *Pers. Soc. Psychol. Rev.* 2:100–10
Metcalfe J, Butterfield B, Habeck C, Stern Y. 2012. Neural correlates of people's hypercorrection of their false beliefs. *J. Cogn. Neurosci.* 24:1571–83. **http://www.mitpressjournals.org/doi/abs/10.1162/jocn_a_00228**
Metcalfe J, Casal-Roscum L, Radin A, Friedman D. 2015. On teaching old dogs new tricks. *Psychol. Sci.* 12:1833–42
Metcalfe J, Finn B. 2008. Evidence that judgments of learning are causally related to study choice. *Psychon. Bull. Rev.* 15:174–79
Metcalfe J, Finn B. 2011. People's correction of high confidence errors: Did they know it all along? *J. Exp. Psychol.: Learn. Mem. Cogn.* 37:437–48
Metcalfe J, Finn B. 2012. Hypercorrection of high confidence in children. *Learn. Instr.* 22:253–61
Metcalfe J, Kornell N. 2007. Principles of cognitive science in education: the effects of generation, errors and feedback. *Psychon. Bull. Rev.* 14:225–29
Metcalfe J, Kornell N, Finn B. 2009. Delayed versus immediate feedback in children's and adults' vocabulary learning. *Mem. Cogn.* 37:1077–87
Metcalfe J, Kornell N, Son LK. 2007. A cognitive-science based program to enhance study efficacy in a high- and low-risk setting. *Eur. J. Cogn. Psychol.* 19:743–68
Metcalfe J, Miele DB. 2014. Hypercorrection of high confidence errors: Prior testing both enhances delayed performance and blocks the return of the errors. *J. Appl. Res. Mem. Cogn.* 3:189–97
Mitchell DJ, Russo JE, Pennington, N. 1989. Back to the future: temporal perspective in the explanation of events. *J. Behav. Decis. Mak.* 2:25–39
Moreno R. 2004. Decreasing cognitive load for novice students: effects of explanatory versus corrective feedback on discovery-based multimedia. *Instr. Sci.: Int. J. Learn. Cogn.* 32:99–113
Morrison JE, Meliza LL. 1999. *Foundations of the After Action Review Process*. US Army Res. Inst. Behav. Soc. Sci., Spec. Rep. 42. Alexandria, VA: Inst. Def. Anal.

Nader K, Schafe G, LeDoux JE. 2000. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature* 406:722–26

Paller KA, Kutas M, Mayes AR. 1985. An investigation of neural substrates of memory encoding in man. *Psychophysiology* 22:607

Pashler H, Cepeda NJ, Wixted JT, Rohrer D. 2005. When does feedback facilitate learning of words? *J. Exp. Psychol.: Learn. Mem. Cogn.* 31:3–8

Proudfit GH, Inzlicht M, Mennin DS. 2013. Anxiety and error monitoring: the importance of motivation and emotion. *Front. Hum. Neurosci.* 7:636

Pyc MA, Rawson KA. 2010. Why testing improves memory: mediator effectiveness hypothesis. *Science* 330:335

Richland LE, Kao LS, Kornell N. 2009. The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *J. Exp. Psychol. Appl.* 15:243–57

Roediger HL, Karpicke JD. 2006a. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychol. Sci.* 17:249–55

Roediger HL, Karpicke JD. 2006b. The power of testing memory: basic research and implications for educational practice. *Perspect. Psychol. Sci.* 1:181–210

Rumelhart DE, McClelland JL. 1986. *Parallel Distributed Processing*, Vol. 1: *Explorations in the Microstructure of Cognition.* Cambridge, MA: MIT Press

Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, et al. 2009. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch. Intern. Med.* 169:1881–87

Schiller D, Monfils M-H, Raio CM, Johnson DC, LeDoux JE, Phelps EA. 2010. Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature* 463:49–53

Scholar AA, Stroessner SJ, Higgins ET. 2008. Responding to negativity: how a risky tactic can serve a vigilant strategy. *J. Exp. Soc. Psychol.* 44:767–74

Siegler RS. 1995. How does change occur: a microgenetic study of number conservation. *Cogn. Psychol.* 28:225–73

Siegler RS. 2002. Microgenetic studies of self-explanation. In *Microdevelopment: A Process-Oriented Perspective for Studying Development and Learning*, ed. N Garnott, J Parziale, pp. 31–58. Cambridge, UK: Cambridge Univ. Press

Sitzman DM, Rhodes MG. 2010. *Does the hypercorrection effect occur when feedback is delayed?* Poster presented at 2010 Psychon. Soc. Annu. Meet., St. Louis, MO

Sitzman DM, Rhodes MG, Tauber SK, Liceralde VR. 2015. The role of prior knowledge in error correction for younger and older adults. *Aging Neuropsychol. Cogn.* 22:502–16

Skinner BF. 1953. *Science and Human Behavior*. New York: MacMillan

Slamecka NJ, Fevreiski J. 1983. The generation effect when generation fails. *J. Verb. Learn. Verb. Behav.* 22:153–63

Slamecka NJ, Graf P. 1978. The generation effect: delineation of a phenomenon. *J. Exp. Psychol.: Hum. Learn. Mem.* 4:592–604

Smith SM, Blankenship SE. 1991. Incubation and the persistence of fixation in problem solving. *Am. J. Psychol.* 104:61–87

Stevenson H, Stigler JW. 1994. *The Learning Gap: Why Our Schools Are Failing and What We Can Learn from Japanese and Chinese Education*. New York: Simon & Schuster

Stigler JW, Hiebert J. 2009. Closing the teaching gap. *Kappan* 91:32–37

Terrace HS. 1963. Discrimination learning with and without errors. *J. Exp. Anal. Behav.* 6:1–27

Terrace HS. 1966. Stimulus control. In *Operant Behavior: Areas of Research and Application*, ed. WK Honig, pp. 271–334. New York: Appleton-Century-Crofts

Terrace HS. 1968. Discrimination learning the peak shift and behavioral contrast. *J. Exp. Anal. Behav.* 11:727–41

Terrace HS. 2010. Defining the stimulus—a memoir. *Behav. Process.* 83:139–53

USAID (US Agency Int. Dev.). 2006. *After-action review. Technical guidance.* PN-ADF-360. Washington, DC: USAID

van Loon MH, Dunlosky J, van Gog T, van Merrienboer, JJG, de Bruin ABH. 2015. Refutations in science texts leads to hypercorrections of misconceptions held with high confidence. *Contemp. Educ. Psychol.* 42:39–48

Wahlheim CN, Jacoby LL. 2013. Remembering change: the critical role of recursive remindings in proactive memory. *Mem. Cogn.* 41:1–15

Waterman AD, Garbutt J, Hazel E, Dunagan WC, Levinson W, et al. 2007. The emotional impact of medical errors on practicing physicians in the United States and Canada. *Jt. Comm. J. Qual. Patient Saf.* 33:467–76

Wiswede D, Münte TF, Goschke T, Rüsseler J. 2009. Modulation of the error-related negativity by induction of short-term negative affect. *Neuropsychology* 47:83–90

Wu AW. 2000. Medical error: the second victim. *BMJ* 320:726–27

Zhao B. 2011. Learning from errors: the role of context, emotion, and personality. *J. Org. Behav.* 32:435–63