Journal of Experimental Psychology: General
2017, Vol. 146, No. 5, 615–620

0096-3445/17/$12.00 http://dx.doi.org/10.1037/xge0000274

BRIEF REPORT

Practice Increases Procedural Errors After Task Interruption

Erik M. Altmann and David Z. Hambrick
Michigan State University

Positive effects of practice are ubiquitous in human performance, but a finding from memory research suggests that negative effects are possible also. The finding is that memory for items on a list depends on the time interval between item presentations. This finding predicts a negative effect of practice on procedural performance under conditions of task interruption. As steps of a procedure are performed more quickly, memory for past performance should become less accurate, increasing the rate of skipped or repeated steps after an interruption. We found this effect, with practice generally improving speed and accuracy, but impairing accuracy after interruptions. The results show that positive effects of practice can interact with architectural constraints on episodic memory to have negative effects on performance. In practical terms, the results suggest that practice can be a risk factor for procedural errors in task environments with a high incidence of task interruption.

Keywords: list memory, temporal distinctiveness, practice effects, procedural error, task interruption

Supplemental materials: http://dx.doi.org/10.1037/xge0000274.supp

Positive effects of practice are ubiquitous in human behavior (see, e.g., Newell & Rosenbloom, 1981), but a finding from memory research suggests that negative effects are possible also. This finding, often characterized as the *ratio rule* (Bjork & Whitten, 1974; Brown, Neath, & Chater, 2007), relates the accuracy of memory for items on a list to the ratio of two quantities: the interpresentation interval (IPI) between onsets of successive items, and the retention interval (RI) between the last item and the memory test. A smaller IPI/RI ratio means that items are represented with less temporal distinctiveness in memory and thus are more confusable. Metaphorically, items become less distinct memorially as they recede into the past much as telephone poles become less distinct visually as they recede into the distance from the vantage point of a moving train (Crowder, 1976, p. 462).

The ratio rule predicts a negative effect of practice on procedural performance under conditions of task interruption. Often, if someone is interrupted while performing a procedure, they have to remember which steps of the procedure they performed before the interruption in order to avoid repeating or skipping steps after the interruption. Mapped to the ratio rule, the time per step for the steps leading up to an interruption represents an IPI between performed steps, and the interruption represents an RI between the last of those steps and a memory test. If practice causes this IPI to decrease relative to this RI, the change should impair memory for steps performed before the interruption and thus accuracy in resuming the procedure afterward.

This prediction has practical significance because many real-world task environments are characterized by highly practiced procedural performance, frequent task interruptions, and costly errors. For example, suppose a nurse is interrupted while preparing to administer a dose of medication, and after the interruption must remember whether he or she administered the dose. The ratio rule predicts that a more practiced nurse should remember less accurately than a less practiced nurse, other things being equal, because the more practiced nurse performs the steps involved in administering medication more quickly. This is not to say that skilled nurses should avoid administering medication, but only that high levels of skill could be a risk factor for increased error after interruptions and that accurate performance may involve compensatory adaptation of some kind.

We tested this prediction of the ratio rule using a task designed to study procedural error under conditions of task interruption (Altmann, Trafton, & Hambrick, 2014). In this task, participants perform a procedure in a continuous loop and are periodically interrupted between steps. After each interruption they have to remember the last step they performed before the interruption in order to resume the procedure correctly. In terms of the ratio rule, the preinterruption step is the target item to be retrieved after the interruption, the response time (RT) per step is the IPI, and the interruption is the RI. The ratio

This article was published Online First March 16, 2017.

Erik M. Altmann and David Z. Hambrick, Department of Psychology, Michigan State University.

This research was supported by grants from the Office of Naval Research (N000140910093 to Erik M. Altmann and N000141310247, N000141612457, and N000141612841 to Erik M. Altmann and David Z. Hambrick). The results were presented at the 2016 Annual Meeting of the Psychonomic Society, Boston, MA.

Correspondence concerning this article should be addressed to Erik M. Altmann, Psychology Building, 316 Physics Road, East Lansing, MI 48824. E-mail: ema@msu.edu

rule predicts that if practice causes RT to decrease relative to interruption duration, errors in resuming the procedure after interruptions should increase.

Method

Participants

Participants were Michigan State University undergraduates. A total of 224 participants completed both sessions of the task.[1] We discarded the data of 18 participants because their accuracy results suggested they were not following instructions, as discussed in the Procedure section. Thus, we analyzed data from 206 participants.

Materials

Figure 1 shows a schematic representation of the task. There is a procedure with seven steps, each identified by a letter that stands for a different two-alternative forced-choice task. The letters spell *UNRAVEL*, an acronym that defines the correct order of the steps.

Participants perform the steps in order, one step per trial, starting over with U when they reach L. On each trial a randomly constructed stimulus is presented that affords performance of any of the seven steps, so participants must remember the last step they performed in order to select the correct step to perform next. Figure 2a shows two sample stimuli. Figure 2b shows the seven choice rules, the two candidate responses for each rule, and the correct response under each rule for each of the two sample stimuli.

Performance of the procedure is interrupted randomly between one trial and the next by a simple typing task. The participant must correctly type two consecutive strings of letters, each of which is a random permutation of the 14 candidate UNRAVEL responses. Figure 2c shows a sample interruption stimulus and partially typed string.

Procedure

Participants performed two sessions of UNRAVEL about 4 days apart (M = 4.31, SD = 1.50). In Session 1, participants received instruction on the task and then performed four experimental blocks. Each block comprised 11 runs of trials, each six trials long on average, separated by 10 interruptions. In Session 2, participants received abbreviated instructions and then performed four experimental blocks identical to those in Session 1 except that each comprised 12 runs of trials separated by 11 interruptions. The first run of each block was treated as a warm-up period and excluded from analysis. Each session took about 30 min.

After each experimental block, participants were given their score for the block, which was the percentage of trials correct on both accuracy measures (defined below). If the score was over 90% the participant was asked to go faster; if the score was under 70% the participant was asked to be more accurate and that block was excluded from analysis. If, in either session, a participant's score was under 70% on two more blocks, or their step selection accuracy was not above chance on postinterruption trials, that individual's data were discarded on grounds that he or she was not following instructions (18 participants).

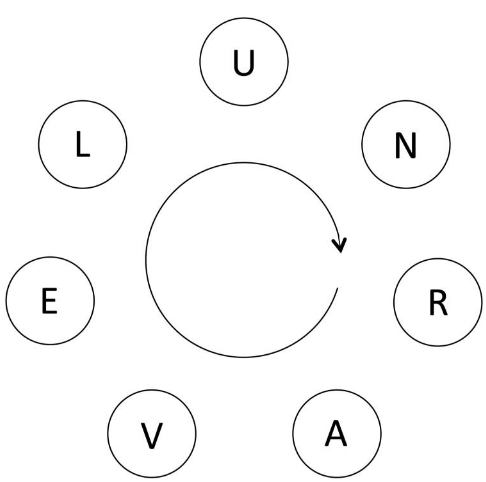


Figure 1. Schematic representation of the behavioral task. Each circle designates a step in the procedure, and the letter in the circle mnemonically identifies the choice rule for that step (see Figure 2b). Participants perform the steps in the order specified by the UNRAVEL acronym, one step per trial, continuing with U after they perform L.

Dependent Variables

We collected two accuracy measures: *sequence errors*, which depend on memory for the previous trial, and *nonsequence errors*, which do not. A sequence error occurs when the participant selects a step that is not the correct successor to the step performed on the previous trial. For example, if a participant performs the U, R, and A steps on three successive trials, a sequence error occurs on the R trial (the correct step would have been N) but not on the A trial (A is the correct successor to R). A nonsequence error occurs when the participant selects the correct step but the incorrect response for that step given the stimulus.

We also collected two timing measures, RT on UNRAVEL trials and the duration of interruptions. We aggregated each measure within participants by taking untrimmed means (use of medians does not change the pattern of results). Trials with sequence errors or nonsequence errors were excluded from RT analysis, as they were analyzed separately as our error data.[2]

Experimental Design

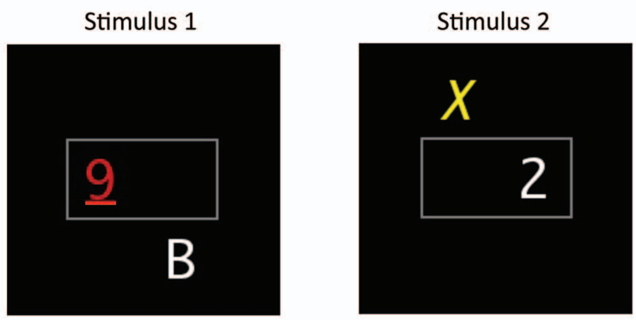
There were three experimental factors, all within participants. The Session factor was discussed above. The Position factor, with Levels 1 through 6, is the serial position of a trial following an interruption. We generally refer to Position 1 as the *postinterruption trial* and Positions 2 through 6 as *baseline trials*. We limit our analysis to the first six positions because relatively few runs of trials are longer than that (Altmann & Trafton, 2015).

The *offset* factor applies to sequence errors only. The levels are −3, −2, −1, +1, +2, and +3, each representing a different number of steps skipped backward (−) or forward (+) in the

[1] Of our total sample of 224 participants, 158 constituted the total sample reported by Hambrick and Altmann (2015). The 66 participants making up the difference are those who completed the two sessions of the task we report here but did not complete all the tasks reported by Hambrick and Altmann, which were distributed across three sessions.

[2] The aggregation methods and exclusion criteria reported here are identical to those reported in all previous publications involving this task (Altmann & Trafton, 2015; Altmann et al., 2014; Altmann, Trafton, & Hambrick, in press; Hambrick & Altmann, 2015).

(a) Two sample UNRAVEL stimuli (the 9 is red and the *X* is yellow):



(b) The choice rule for each UNRAVEL step, the candidate responses for each rule, and for each rule the correct response to each stimulus in (a):

Step	Choice rule	Candidate responses		Stimulus 1 response	Stimulus 2 response
U	character is Underlined or in Italics	u	i	u	i
N	letter is Near to or Far from start of alphabet	n	f	n	f
R	character is Red or Yellow	r	y	r	y
A	character is Above or Below the box	a	b	b	a
V	letter is Vowel or Consonant	v	c	c	c
E	digit is Even or Odd	e	o	o	e
L	digit is Less than or More than 5	l	m	m	l

(c) Sample interruption stimulus, after the participant has typed four letters of the first of two strings ("code #1"):

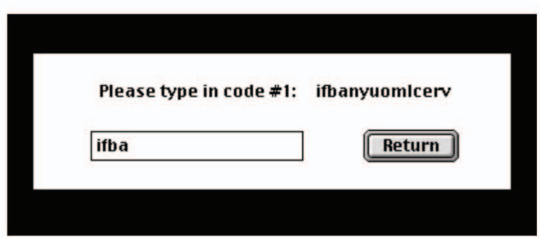


Figure 2. Stimuli and response rules for the behavioral task. See the online article for the color version of this figure.

UNRAVEL sequence. For example, if someone performs the R step on one trial, the correct successor is A. If the step performed instead is U, N, R, V, E, or L, the offset of that sequence error is −3, −2, −1, +1, +2, or +3, respectively.

Results

If the ratio of RT to interruption duration decreased between sessions, then postinterruption sequence errors should have increased between sessions. The ratio did in fact decrease, from .105 to .092, $F(1, 205) = 122.70$, $p < .001$, $\eta_p^2 = .374$,[3] and postinterruption sequence errors did in fact increase, from .124 to .142, $F(1, 205) = 5.43$, $p = .021$, $\eta_p^2 = .026$. Sequence errors are plotted in the top panel of Figure 3.

All other error measures decreased between sessions. Sequence errors decreased on baseline trials, $F(1, 205) = 6.02$, $p = .015$, $\eta_p^2 = .029$. Nonsequence errors, plotted in the middle panel of Figure 3, decreased on the postinterruption trial, $F(1, 205) = 14.28$, $p < .001$, $\eta_p^2 = .065$, and on baseline trials, $F(1, 205) = 12.29$, $p = .001$, $\eta_p^2 = .057$.

[3] The RT measure we used to compute the ratio was the mean across baseline trials, which decreased from 2.65 s in Session 1 to 2.14 s in Session 2, $F(1, 205) = 523.40$, $p < .001$, $\eta_p^2 = .719$. Interruption duration decreased between sessions also, from 22.31 s to 20.66 s, $F(1, 205) = 74.12$, $p < .001$, $\eta_p^2 = .266$, but the decrease was smaller proportionally than that for RT, so the decrease in RT dominated the effect on the ratio.

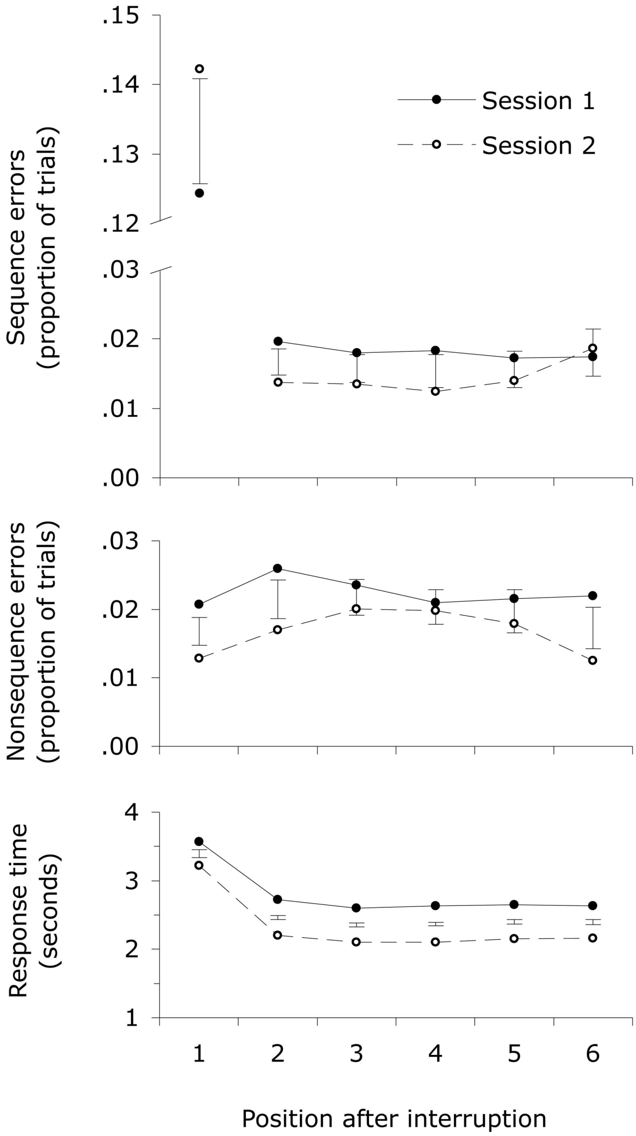


Figure 3. Behavioral data from the experiment. Error bars are graphical significance tests (Altmann et al., 2014, note 2) for the effect of session at each position. The effect is significant if the markers lie outside the bar.

RT, plotted in the bottom panel of Figure 3, also became faster between sessions, $F(1, 205) = 418.02$, $p < .001$, $\eta_p^2 = .671$. However, the effect differed by position, $F(5, 1025) = 4.52$, $p < .001$, $\eta_p^2 = .022$, with less speed-up on the postinterruption trial than on baseline trials. RT was also slower on the postinterruption trial than on baseline trials, as reflected in a main effect of position, $F(5, 1025) = 191.90$, $p < .001$, $\eta_p^2 = .483$. In terms of subtractive logic, this pattern of effects is evidence for a processing stage unique to the postinterruption trial that became slower between sessions. We interpret this stage as retrieval of memory for the preinterruption step.

Sequence errors are replotted in Figure 4 separated by the offset factor. This view shows that the increase in postinterruption sequence errors (upper left panel) was driven by errors at Offset −1, which are repetitions of the preinterruption step. An omnibus analysis of variance (see Table S2 in the online supplemental materials) revealed a Session × Position × Offset interaction, $F(25, 5125) = 2.73$, $p < .001$, $\eta_p^2 = .013$, and follow-up analyses on the postinterruption trial showed a Session × Offset interaction, $F(5, 1025) = 3.99$, $p = .001$, $\eta_p^2 = .019$, a session effect at Offset −1, $F(1, 205) = 9.72$, $p = .002$, $\eta_p^2 = .045$, and no session effect across the other offsets ($F < 1$).

Theoretical Model

An existing model of performance on our task (Altmann & Trafton, 2015; Altmann et al., in press) implements the ratio rule in a way that accounts for the increase in postinterruption sequence errors specifically at Offset −1 (compare the upper panels of Figure 4). The model incorporates a representation of the local distinctiveness principle (e.g., Brown et al., 2007), which states that the distractor most likely to intrude on a target is the target's nearest neighbor.[4] In our task the target is a memory for the preinterruption trial, so the nearest neighbor of the target is a memory for the pre-preinterruption trial. If this nearest neighbor intrudes, the system will select the preinterruption step to perform next, because the system infers that the step to perform next is the successor of the remembered step. The nearest neighbor was temporally nearer to the target in Session 2 than in Session 1, because RT was faster in Session 2. Therefore, the preinterruption step was incorrectly selected more often in Session 2 than in Session 1, driving the session effect on errors at Offset −1. The gradient of local distinctiveness, which is constrained by the underlying theoretical function, correctly predicts that neighbors more distant than the pre-preinterruption trial were too distant to have much influence on performance, limiting the session effect to Offset −1. Quantitatively, the session effect on baseline RT fully accounts for the session effect on postinterruption sequence errors, with no residual variance to be explained and no free parameters to be adjusted (as we discuss in the online supplemental materials).

Our model does not account for the decrease in baseline sequence errors between sessions (compare the lower panels of Figure 4). We assume that the system has to remember the last performed step on baseline trials just as it does on the postinterruption trial, and we estimate the RI for this retrieval as the RT, representing time elapsed since the last trial. Thus, the ratio of RT to RI is RT/RT, a constant. This analysis may oversimplify the effects of practice on baseline performance, as we discuss below.

Discussion

We found that practice impaired the ability to resume a procedural task after interruptions. Sequence errors increased on the postinterruption trial, as did the duration of a processing stage unique to that trial. The two effects together suggest that recall of the last step performed before the interruption became less accurate and slower. All other measures of performance improved with practice, including RT on baseline trials. Our model links this improved RT to reduced temporal distinctiveness of memory for

[4] The model actually incorporates a representation of decay theory, which is generally viewed in opposition to distinctiveness theory, but the two turn out to be identical for the case in which the retrieval target is the most recent item, as we show in the online supplemental materials.

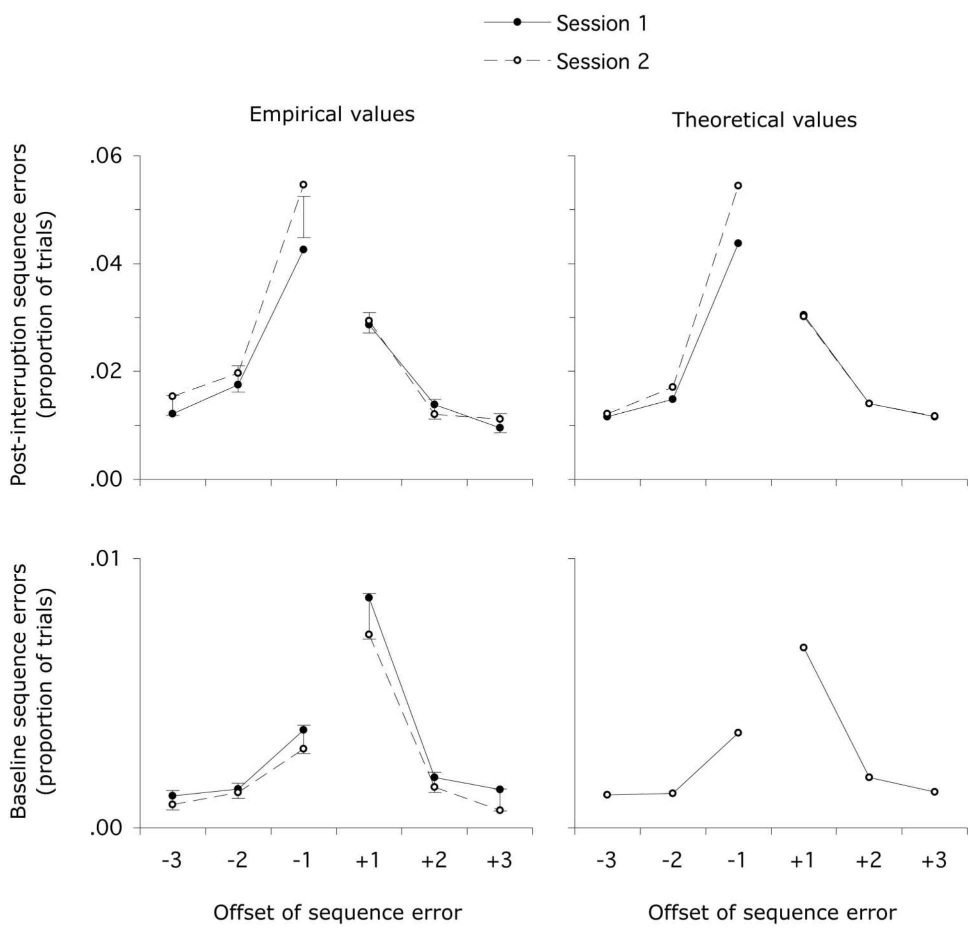


Figure 4. Sequence error data from the experiment (left panels) and corresponding theoretical values from the model described in the text (right panels; in the lower-right panel, the two series are on top of one another). The postinterruption trial is Position 1 after an interruption, and baseline trials are Positions 2 through 6 after an interruption. Error bars are graphical significance tests (Altmann et al., 2014, note 2) for the effect of session at each offset. The effect is significant if the markers lie outside the bar.

preinterruption events, making them more confusable. The gradient of temporal distinctiveness in our model predicts that only the nearest neighbor to the target is likely to intrude, explaining why the increase in sequence errors was driven by repetitions of the preinterruption step (i.e., errors at Offset −1; see Figure 4).

At least two other mechanisms may have affected our results. First, proactive interference in memory for performed steps could have been higher in Session 2 than in Session 1. This mechanism would explain the increase in postinterruption sequence errors between sessions, but not the increase specifically at Offset −1, as all performed steps would have had an increased likelihood of intruding. Second, practice could have strengthened the associations between steps, making baseline performance less reliant upon explicit memories for performed steps. This mechanism would explain why baseline performance became faster and more accurate between sessions. If, in addition, explicit memories were encoded less reliably, this mechanism would explain the increase in postinterruption sequence errors between sessions—but again would not seem to explain the error increase specifically at Offset −1, as all performed steps would have had the same decreased likelihood of being encoded. That said, these or other mechanisms may fully account for our data when formalized in detail. To facilitate development of alternative models we have made our complete data set available online (see the online supplemental materials).

The session effects in our data resemble a speed–accuracy trade-off, though not the usual kind. In the usual kind, the trading relation is between speed and accuracy on the same kind of trial, and reflects a strategic shift of emphasis from one measure to the other (e.g., Wickelgren, 1977). In our data this tradeoff was absent; baseline performance became faster and more accurate, and memory retrieval on the postinterruption trial became slower and less accurate. Instead, the trading relation in our data was between speed and accuracy on different kinds of trials (baseline vs. postinterruption), and represented not a shift in strategy, according to our model, but a negative effect of practice interacting with operating principles of human memory.

In practical terms, our results suggest that training can be a risk factor for increased procedural error in task environments with a high incidence of interruption. This is not to suggest that training is counterproductive, but rather that procedural performance that is fast and accurate despite frequent interruption may be based in part on adaptations that help with placekeeping—possibly representing an opportunity to introduce relevant strategies during training, or include features that support placekeeping in the design of the task environment.

Compressed memory for past events is not the only mechanism that could mediate negative effects of practice on accuracy. Reason's (1990) influential work on errors suggests two others, both linking elevated error rates to changes in routine. Practice could increase the strength of default steps or actions, increasing the error rate at choice points where the default happens to be incorrect. Practice could also decrease the frequency of "attentional checks" on progress toward the goal, increasing the error rate when conditions dictate a change of course. Reason's analysis focused on errors in diary studies and occupational case studies, which have high external validity but offer limited constraint on models of underlying mechanisms. Developing procedures to study these mechanisms in the laboratory is an important challenge for future work.

References

Altmann, E. M., & Trafton, J. G. (2015). Brief lags in interrupted sequential performance: Evaluating a model and model evaluation method. *International Journal of Human–Computer Studies, 79,* 51–65. http://dx.doi.org/10.1016/j.ijhcs.2014.12.007

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2014). Momentary interruptions can derail the train of thought. *Journal of Experimental Psychology: General, 143,* 215–226. http://dx.doi.org/10.1037/a0030986

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (in press). Effects of interruption length on procedural errors. *Journal of Experimental Psychology: Applied.*

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6,* 173–189. http://dx.doi.org/10.1016/0010-0285(74)90009-7

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539–576. http://dx.doi.org/10.1037/0033-295X.114.3.539

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Hambrick, D. Z., & Altmann, E. M. (2015). The role of placekeeping ability in fluid intelligence. *Psychonomic Bulletin & Review, 22,* 1104–1110. http://dx.doi.org/10.3758/s13423-014-0764-5

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Reason, J. (1990). *Human error*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139062367

Wickelgren, W. A. (1977). Speed–accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85. http://dx.doi.org/10.1016/0001-6918(77)90012-9

Received July 20, 2016
Revision received December 20, 2016
Accepted December 21, 2016 ■