Journal of Experimental Psychology: Applied
2017, Vol. 23, No. 2, 216–229

1076-898X/17/$12.00 http://dx.doi.org/10.1037/xap0000117

Effects of Interruption Length on Procedural Errors

Erik M. Altmann
Michigan State University

J. Gregory Trafton
Naval Research Laboratory, Washington, DC

David Z. Hambrick
Michigan State University

We investigated effects of task interruption on procedural performance, focusing on the effect of interruption length on the rates of different categories of error at the point of task resumption. Interruption length affected errors involving loss of place in the procedure (*sequence errors*) but not errors involving incorrect execution of a correct step (*nonsequence errors*), implicating memory for past performance, rather than generalized attentional resources, as the disrupted cognitive process. Within the category of sequence errors, interruption length produced a complex pattern of effects, with repetitions of the preinterruption step showing different effects than errors at other offsets from the correct step. A cognitive model we developed previously accounts for the results in terms of decay and rehearsal of memory for past performance and activation spreading through a procedural representation of task knowledge. The model links different types of errors to different cognitive processes, informs potential interventions, and predicts interruption effects for sequential tasks like problem solving and counting.

Keywords: memory, procedural error, task interruption, cognitive control

One important characteristic of many tasks that people encounter at work or in the home is that they are procedural, meaning that they involve a sequence of steps that must be performed in a specific order without omissions or repetitions. Consider the soldier's or police officer's task of cleaning a weapon. An important step is to check that the chamber is empty, and skipping this step is a potentially catastrophic omission. Alternatively, consider the task of administering a dose of medication to a patient. Skipping the dose would cause one problem, but administering the dose and failing to record it on the patient's chart could cause a different problem, if someone then repeats the dose later because there is no record that it was already administered. Many tasks in the medical domain are vulnerable to these kinds of errors, as are tasks in many other domains, including maintenance procedures, legal procedures, computer programming and technical support, building construction and civil engineering, data analysis, tax preparation and accounting, chores around the home, and so on. Sequential constraints also play a role in such basic cognitive processes as serial recall, language production (Dell, Burger, & Svec, 1997), event counting (Carlson & Cassenti, 2004), and problem solving (Carpenter, Just, & Shell, 1990).

A second important characteristic of many of the same task environments is that performance is susceptible to interruptions of various kinds and lengths. Interruptions take many forms, including communications from others (e-mails and texts, phone calls, knocks on the door) or internally generated task switches (Katidioti, Borst, & Taatgen, 2014). When an interruption is over—assuming one remembers there is an interrupted task to return to (Grundgeiger, Sanderson, & Dismukes, 2014)—there may ensue a "Where was I?" moment as one tries to recall one's exact place in the task. Such moments seem to be commonplace in everyday experience. In laboratory tasks, interruptions generally cause a time lag (e.g., Altmann & Trafton, 2007; Hodgetts & Jones, 2006; Monk, Boehm-Davis, & Trafton, 2004) or an increase in error rates (e.g., Altmann, Trafton, & Hambrick, 2014; Li, Blandford, Cairns, & Young, 2008) at the point of task resumption. In the field, interruptions increased medication administration errors by over 12% in one study (Westbrook, Woods, Rob, Dunsmuir, & Day, 2010).

Here we examine the effects of the length of an interruption on performance immediately after the interruption, focusing on the empirical question of whether different kinds of errors are affected differently by interruption length, and the related theoretical question of what underlying cognitive mechanisms mediate the effects. Concerning the empirical question, the task we use (Altmann et al., 2014) affords measurement of various kinds of errors. Sequence errors reflect loss of place in the procedure, whereas nonsequence errors reflect incorrect execution of the correct step. Further, a sequence error can be near to or far from the correct step within the task sequence. Dissociations between these different kinds and levels of error would have practical implications for task environments in which some errors are costlier than others.

This article was published Online First February 2, 2017.

Erik M. Altmann, Department of Psychology, Michigan State University; J. Gregory Trafton, Naval Research Laboratory, Washington, D.C.; David Z. Hambrick, Department of Psychology, Michigan State University.

This research was supported by grants from the Office of Naval Research, N000140910093 to the first author; N000141310247, N000141612457, and N000141612841 to the first and third authors; and N0001412RX20082 and N0001411WX30014 to the second author.

Correspondence concerning this article should be addressed to Erik M. Altmann, Michigan State University, Psychology Building, 316 Physics Road, Room 298A, East Lansing, MI 48824. E-mail: ema@msu.edu

Concerning the theoretical question, we recently developed a model of the memory mechanisms involved in selecting the next step of a procedure under conditions of task interruption (Altmann & Trafton, 2015). The model has broad scope, incorporating representations of procedural knowledge and how it is primed during performance, episodic memory for past performance and how it decays over time, and rehearsal of placekeeping information during interruptions. We use this model here to account for the effects of our interruption length manipulation specifically on sequence errors.

For the interruption length manipulation itself we chose levels that map onto a range of situations that people encounter outside the lab. Our shortest length is about 3 s, which is roughly the time required to find one's phone to turn it off when it starts ringing in the middle of a meeting. Our longest length is about 30 s, which is long enough to have a meaningful conversation if one chooses to answer the phone, if only to negotiate a better time to talk. The manipulation is parametric, with four levels, to help isolate the nature of interactions between interruption length and type of error that might have practical or theoretical implications.

To preview our results, we found that interruption length affected sequence errors but not nonsequence errors, suggesting that disrupted memory was central to the behavioral effects of interruption. The effect on sequence errors resembled a standard curvilinear forgetting function (Rubin & Wenzel, 1996), but only at an aggregate level. At a more detailed level, the effect varied with the offset of the error, meaning the proximity of the incorrect step to what would have been the correct step. Specifically, repetitions of the most recently performed step showed a markedly different trend across levels of interruption length than other sequence errors. We examine the theoretical and practical implications of this pattern in the General Discussion.

Method

Participants

Participants were members of the Michigan State University community. The sample size was 400, with 100 participants included in each of four interruption length conditions (very short, short, medium, and long). Participants received either credit toward a course requirement or payment of $10 (none were paid in the very short condition, 10 were paid in the short condition, 18 were paid in the medium condition, and 16 were paid in the long condition). Sixteen additional participants were excluded because they performed below a target accuracy level that we describe in the Procedure section. Data from one condition were previously published (Altmann et al., 2014), as we describe in the Experimental Design section.

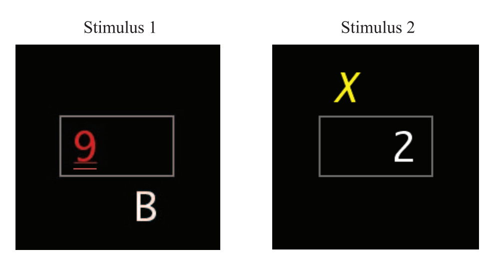
Materials

The procedural component of the task is defined by the acronymic word UNRAVEL. Each letter identifies a step of the procedure, and the letter sequence identifies the correct order of the steps. The order is therefore defined in terms of a word likely to be in the lexicon of English-speaking participants, so that interpretation of sequence errors is not confounded by lack of task knowledge on the part of the performer (Reason, 1990). Participants perform the UNRAVEL sequence in a loop, starting over with U after they reach L. In an average session the participant cycles through the loop about 38 times.

On each trial the participant performs one step. A stimulus is presented and the participant makes a choice regarding the stimulus feature dictated by the step the participant thinks is correct for that trial. Figure 1a shows two sample stimuli that illustrate the various stimulus features. Each stimulus has two characters, one a letter (A, B, U, or X) and one a digit (1, 2, 8, or 9), both randomly selected from the set of four options. Each stimulus also has a font style (underline or italics), color (red or yellow), and location outside the gray outline box (above or below), each randomly selected from the set of two options, and then assigned randomly and independently to one character or the other.

Figure 1b shows the choice rule for each step. The choice for the U step is whether the font style is underline or italic, for the N step is whether the letter is near to or far from the start of the alphabet, for the R step is whether the color is red or yellow, for the A step is whether the character outside the box is above or below, for the V step is whether the letter is a vowel or a consonant, for the E step is whether the digit is even or odd, and for the L step is whether the digit is less than or more than 5. The letter for each step mnemonically identifies one of the two candidate responses for that step—*u* for underline, *n* for near to, *r* for red, *a* for above, *v* for vowel, *e* for even, and *l* for less-than. All 14 of the candidate responses are different, so any given response indicates what step the participant thought was correct on that trial. Each stimulus

a



b

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

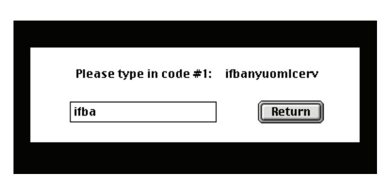


Figure 1. (a): Two sample stimuli for the UNRAVEL task (the 9 is red and the *X* is yellow). (b): Choice rules and candidate responses for the UNRAVEL task, and the correct responses according to each rule for the two sample stimuli in (a). (c): A sample display for the interrupting task, after the participant has typed the first four letters of the "code." See the online article for the color version of this figure.

affords the performance of any of the seven steps, so information about which step to perform next has to be maintained in memory. There is no particular significance to the number of rules other than the practical constraints involved in developing an acronym that includes a candidate response from every rule.

Performance is interrupted periodically between one trial and the next, on average every six trials. The number of trials between interruptions is three or greater, with the exact number randomized to produce a flat hazard function, after the third trial, for the next trial being the last trial before the next interruption (Altmann et al., 2014). The average of six trials between interruptions, given the procedure length of seven steps, ensures that interruptions are evenly distributed across all steps. An interruption begins immediately after a response to a trial, with presentation of a string of letters to type into a box. Figure 1c shows a sample interruption display. The letters to be typed are sampled from the 14 candidate UNRAVEL responses, our aim being to frustrate use of the keyboard as an external memory for placekeeping information. After typing the letters, the participant presses the Return key, and the computer detects any errors at that point. If there are errors, the screen flashes, the box is cleared, and the participant tries to type the same string again. We manipulated interruption length by varying the number of letters to be typed, as described in the Experimental Design section.

Procedure

Participants were tested individually in sessions lasting 30 to 50 min, with much of the variability in session length due to the differences in interruption length. A session began with an introduction to the UNRAVEL sequence that described each step and emphasized how it corresponded mnemonically to a letter of the acronym. After all the steps were introduced, a screen appeared spelling out the acronym and showing the choice rule for each letter (essentially the first two columns of Figure 1b). After this, there were 16 practice trials during which the computer required the correct response on each trial before allowing the participant to move on. This 16-trial sequence was interrupted twice to illustrate the requirement to resume one's place in the sequence after an interruption. The experimenter remained present during this period to help if necessary. A sheet of paper with the information in the first two columns of Figure 1b remained visible to the side of the computer throughout the session.

In preparation for the test phase of the session, participants were reminded to "please try to keep your place in the UNRAVEL sequence," and, after an interruption, to "please try to pick up in the sequence where you left off." During the test phase the computer accepted any of the 14 candidate responses in Figure 1b as the response for any trial. A sequence error was coded as a step performed on one trial that was not the immediate successor in UNRAVEL of the step performed on the previous trial. For example, if a participant performed U, R, and A on three successive trials, the R trial was coded as a sequence error because N was skipped, but the A trial was coded as a correct step because the previous trial was R. A nonsequence error was defined as an incorrect choice for the correct step. No error feedback was given after trials. After a response, the next event (trial or interruption) began immediately.

There were four blocks of trials in the test phase, each with 10 interruptions. There were six trials between interruptions, on average, as we described earlier, and there was a run of trials before the first interruption, so there were approximately $11 \times 6 = 66$ trials per block and $4 \times 66 = 264$ trials per session. After each block the computer presented the participant's score, computed as the percentage of trials that block with neither a sequence error nor a nonsequence error. If the score was above 90%, the participant was asked to go faster; if the score was below 70%, the participant was asked to be more accurate and that block was excluded from analysis (17 cases: one in the very short condition, three in the short condition, nine in the medium condition, and four in the long condition; conditions are defined in the Experimental Design section). A participant was replaced if they scored below 70% on two or more blocks (12 cases: four in the very short condition, two in the short condition, and six in the long condition). A participant was also replaced if their accuracy on the postinterruption trial was not significantly above chance (four additional cases: one in the short condition, one in the medium condition, and two in the long condition); in such cases, we assumed the participant was not following the instruction to try to resume their place in the sequence after an interruption.

Experimental Design

There were three experimental factors: interruption length, offset, and trial type. *Interruption length* is a between-participants factor with four levels, differing in the number of letters to be typed during an interruption. For very short interruptions, participants typed two letters per interruption (net of errors). This condition is previously published data (Experiment 2 of Altmann et al., 2014). For short, medium, and long interruptions, participants typed one, two, or three sets of 14 letters per interruption, respectively, with each set comprising a random permutation of the 14 candidate UNRAVEL responses (Figure 1b). These three conditions are new data. The procedure for the four conditions was identical except for interruption length, and the samples were drawn from the same population, so we treat them as matched by random assignment. To evaluate the similarity of the groups, we also compare them on measures not affected by interruption length. Interruption durations for the four levels were 2.76 s, 13.12 s, 22.02 s, and 31.91 s. (Interruption durations were means of untrimmed participant means, measured from onset of the interruption display to the Return keypress for the last correctly typed "code.") The spacing between durations was similar enough that for purposes of trend analysis we treat interruption length as a linear scale.

Offset is a within-participant factor that applies only to sequence errors (not to nonsequence errors or response times). The offset of a sequence error is the number of steps skipped forward or backward in the task sequence when that error occurred. In our task, there are seven steps and therefore six offsets: −3, −2, −1, +1, +2, and +3. For example, if a participant performed the R step on one trial and the U, N, R, V, E, or L step on the next trial, the offset of the sequence error on the latter trial would be −3, −2, −1, +1, +2, or +3, respectively.

Trial type is a within-participant factor with two levels, postinterruption and baseline. The postinterruption trial is the first trial after an interruption and bears the brunt of interruption effects in this procedure (Altmann et al., 2014; cf. Altmann & Trafton, 2007). Baseline trials are all other trials and provide a measure of uninterrupted performance to contrast with interrupted performance.

Results

Figure 2 plots sequence errors, nonsequence errors, and response times (RTs) as a function of interruption length and trial type. Sequence errors are aggregated over offset. RTs are means of untrimmed participant means on correct trials, measured from onset of a trial stimulus to the response to that stimulus.

We examined each of the three measures with a 4 (interruption length) × 2 (trial type) analysis of variance (ANOVA). Sequence errors increased with interruption length, $F(3, 396) = 21.94$, $p < .001$, $\eta_p^2 = .143$, and were more frequent on the postinterruption trial than on baseline trials, $F(1, 396) = 365.27$, $p < .001$, $\eta_p^2 = .480$. The Interruption Length × Trial Type interaction was significant, $F(3, 396) = 29.12$, $p < .001$, $\eta_p^2 = .181$, and follow-up analyses revealed an effect of interruption length for the postinterruption trial, $F(3, 396) = 25.86$, $p < .001$, $\eta_p^2 = .164$, but not for baseline trials, $F < 1$.

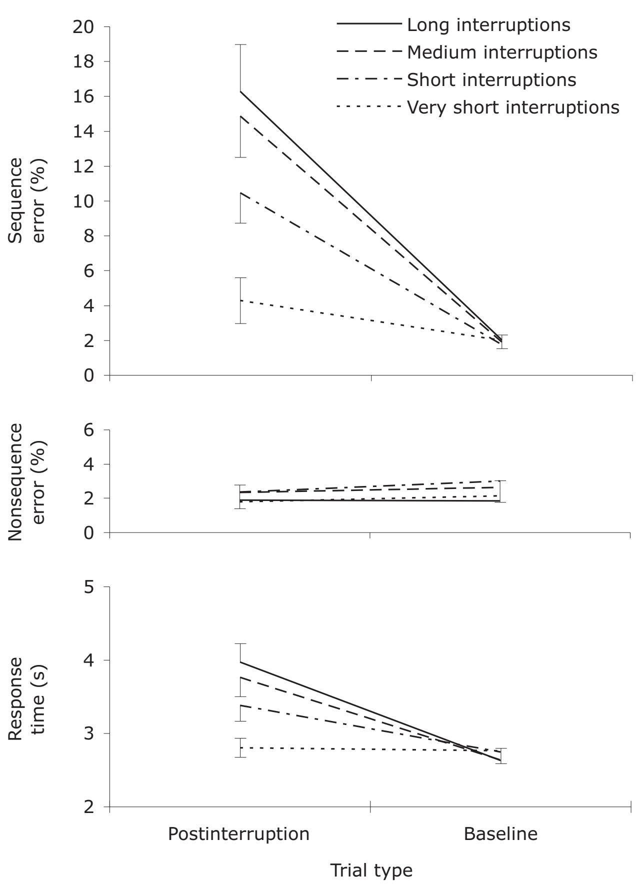


Figure 2. Performance measures on the postinterruption trial and baseline trials. Error bars are 95% confidence intervals (CIs). For baseline trials, the error bar is the average CI across interruption lengths, centered on the mean of the measure across interruption lengths.

Nonsequence errors were not affected by interruption length, $F(3, 396) = 1.67$, $p = .174$, $\eta_p^2 = .012$, but were less frequent on the postinterruption trial ($M = 2.09\%$) than on baseline trials ($M = 2.40\%$), $F(1, 396) = 5.23$, $p = .023$, $\eta_p^2 = .013$. The Interruption Length × Trial Type interaction was not significant, $F = 1.05$.

RT increased with interruption length, $F(3, 396) = 9.59$, $p < .001$, $\eta_p^2 = .068$, and was slower for the postinterruption trial than for baseline trials, $F(1, 396) = 272.44$, $p < .001$, $\eta_p^2 = .408$. The Interruption Length × Trial Type interaction was significant, $F(3, 396) = 37.02$, $p < .001$, $\eta_p^2 = .219$, and follow-up analyses revealed an effect of interruption length for the postinterruption trial, $F(3, 396) = 21.35$, $p < .001$, $\eta_p^2 = .139$, but not for baseline trials, $F(3, 396) = 1.84$, $p = .140$, $\eta_p^2 = .014$. Thus, the RT results generally mirrored the sequence error results.[1]

In Figure 3, the upper panel plots postinterruption sequence errors from Figure 2 against the average duration of the interruption at each level of interruption length, to highlight the trend in error rates across interruption lengths. The trend was negatively accelerating, with trend analysis showing both a significant linear component, $F(1, 396) = 72.43$, $p < .001$, $\eta_p^2 = .155$, and a significant quadratic component, $F(1, 396) = 5.09$, $p = .025$, $\eta_p^2 = .013$.

The lower panel of Figure 3 plots the data from the upper panel separated by the step of the UNRAVEL sequence that would have been correct on the postinterruption trial had there been no sequence error. We found previously that interruption effects were not strongly modulated by this factor, even though the steps have different levels of difficulty (Altmann et al., 2014), and we found a similar pattern here. A 4 (interruption length) × 7 (step in UNRAVEL) ANOVA revealed an effect of interruption length, $F(3, 390) = 26.43$, $p < .001$, $\eta_p^2 = .169$, an effect of step, $F(6, 2340) = 10.54$, $p < .001$, $\eta_p^2 = .026$, and a marginal interaction, $F(18, 2340) = 1.55$, $p = .065$, $\eta_p^2 = .012$. The step effect and the marginal interaction were driven by the U step. A follow-up analysis on the U step revealed an effect of interruption length, $F(3, 390) = 4.67$, $p = .003$, $\eta_p^2 = .035$, a linear trend, $F(1, 390) = 11.97$, $p = .001$, $\eta_p^2 = .030$, and no quadratic trend, $F < 1$. A follow-up analysis on steps other than U revealed an effect of interruption length, $F(3, 390) = 27.16$, $p < .001$, $\eta_p^2 = .173$, a linear trend, $F(1, 390) = 73.78$, $p < .001$, $\eta_p^2 = .159$, and a quadratic trend, $F(1, 390) = 6.33$, $p = .012$, $\eta_p^2 = .016$, but no effect of step, $F(5, 1950) = 1.73$, $p = .126$, $\eta_p^2 = .004$, and no Interruption Length × Step interaction, $F = 1.03$. The lower error rate on U than on other steps suggests that when participants forgot the correct step, they tended to resume with the "first" step of the sequence, causing an elevated rate of guessing the U step correctly.

Figure 4 plots postinterruption sequence errors by offset, revealing a complex pattern that was masked in Figure 3, where the data were aggregated over offset. Specifically, Offset −1 (filled circles) showed different effects of interruption length than did other offsets. (Offsets −2 and +2 showed similar effects, as did Offsets −3 and +3, and these pairs remain aggregated in Figure 4 to

[1] In the Experimental Design section, we noted that the very short condition was previously published, raising the question of whether that group was comparable to the others. The null effects of interruption length on all measures of baseline performance are evidence that the groups were similar and that we are justified in treating the very short condition as one level of the experimental manipulation.

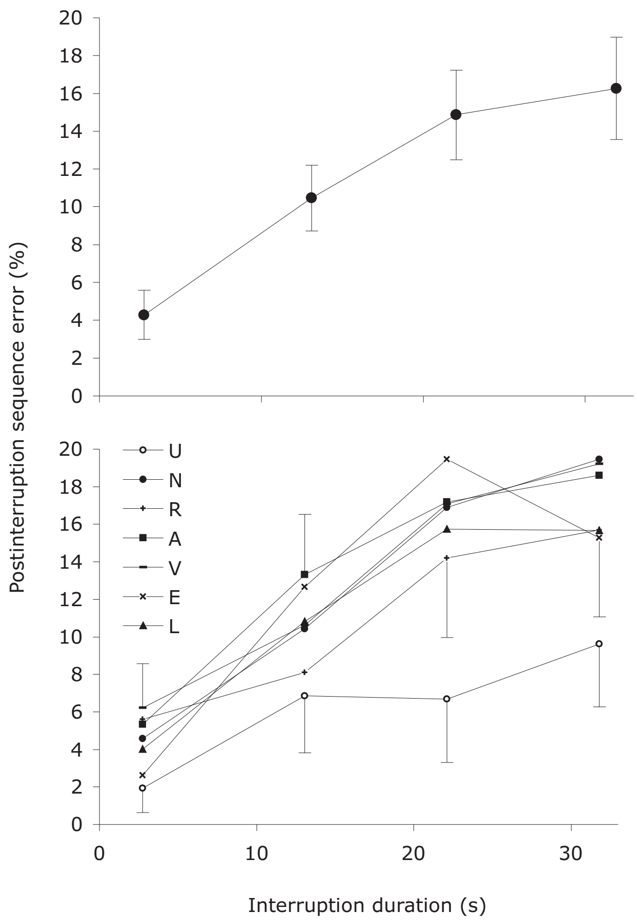


Figure 3. Postinterruption sequence errors. Interruption durations for interruption length conditions: very short = 2.76 s, short = 13.12 s, medium = 22.02 s, and long = 31.91 s. The bottom panel shows the data from the top panel separated by the step in the UNRAVEL sequence that would have been correct. Error bars are 95% confidence intervals (CIs). In the bottom panel, error bars for steps other than U are the average CI across steps other than U.

reduce clutter; all offsets are plotted separately in Figure A2 in the Appendix.) Errors at Offset −1 increased sharply between very short and short interruption lengths and then leveled off, whereas the error rate at other offsets increased gradually and linearly. A 4 (interruption length) × 6 (offset) ANOVA showed a significant Interruption Length × Offset interaction, $F(15, 1980) = 4.83$, $p < .001$, $\eta_p^2 = .035$. One-way ANOVAs on interruption length for each level of offset showed significant main effects and significant linear trends in each case, $ps < .001$. For Offset −1 there was also a significant quadratic trend, $F(1, 396) = 17.32$, $p < .001$, $\eta_p^2 = .042$, but this was largely driven by the difference between very short and short interruptions, with no significant difference across short, medium, and long interruptions, $F(2, 297) = 1.60$, $p = .204$, $\eta_p^2 = .011$. No other offset showed a quadratic trend, $ps > .330$. There were not enough observations per participant to separate the data by step, so we could not extend the analysis in the bottom panel of Figure 3.

Discussion

We examined effects of interruption length on procedural performance parametrically across a range of practically relevant interruption durations—from about 3 s to about 30 s. Without considering the offset factor, which measures the proximity of a sequence error to the correct step, sequence errors on the postinterruption trial increased with interruption duration with the kind of negatively accelerating trend (Figure 3, upper panel) one might expect from a standard forgetting function. Postinterruption RTs mirrored this pattern, with slower responses at longer interruption lengths (Figure 2, lower panel). Nonsequence errors showed no effect of interruption length (Figure 2, middle panel).

Taking the offset factor into account, postinterruption sequence errors showed a complicated pattern (Figure 4) that is difficult to interpret in terms of a single forgetting mechanism. Repetitions of the most recently performed step (i.e., errors at Offset −1) increased sharply from very short to short interruptions, and then leveled off across short, medium, and long interruptions, whereas

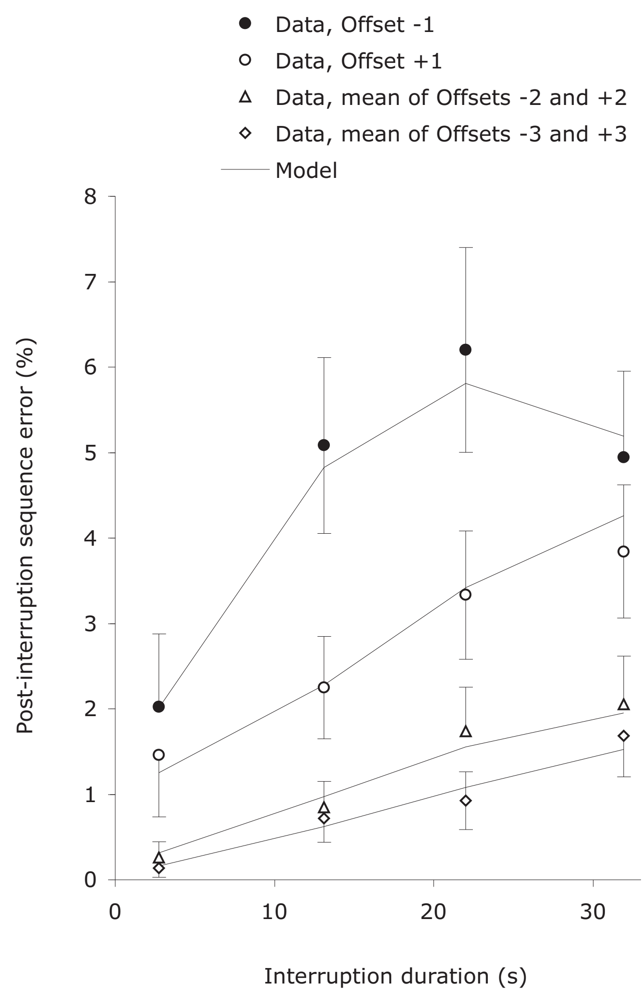


Figure 4. Postinterruption sequence errors by offset. Interruption durations for interruption length conditions: very short = 2.76 s, short = 13.12 s, medium = 22.02 s, and long = 31.91 s. Markers are empirical values, error bars are 95% confidence intervals, and lines are model values.

errors at other offsets increased more gradually and essentially linearly across the full range of interruption lengths.

In the following sections, we interpret the results theoretically, identify practical implications, and address some limitations of our results and method.

Theoretical Interpretation

Our results speak to two different theoretical questions. The first concerns the cognitive basis of the disruptive effects of interruption. Phenomenologically it often seems that interruptions have a startling or discombobulating effect such that after an interruption it can be difficult to focus again on the interrupted task. In theoretical terms one could explain this effect in terms of a generalized loss of attentional resources available to focus on task resumption.

Our data weigh against such an attentional resources account, at least as applied to the kind of interruption conditions we studied here. If attention, broadly defined, were disrupted following an interruption, an effect of interruption length should have been visible on all behavioral measures on the postinterruption trial. Instead, nonsequence errors showed no effect of interruption length, and indeed were less frequent on the postinterruption trial than on baseline trials (Figure 2). The pattern across the three measures suggests that the effect of interruption length on sequence errors and RTs was linked to degraded memory for placekeeping information from before the interruption. One important qualification is that interruptions in our task environment may not have been the best measure of interruptions that are startling or discombobulating, a point we revisit in the Limitations section.

The second theoretical question concerns the nature of the memory representations that degrade during interruptions. As plotted in Figure 3, the effect of interruption length on sequence errors suggests a single memory representation degrading according to a standard curvilinear forgetting function (Rubin & Wenzel, 1996), as we envisioned in our original conception of interruption effects (Altmann & Trafton, 2002). However, when sequence errors are separated by offset, as in Figure 4, the effect of interruption length is more complex. Indeed, the familiar shape of the forgetting function in Figure 3 turns out to be an illusion caused by aggregation.

To account for our results, we fit the data with a model of placekeeping developed by Altmann and Trafton (2015). That study addressed different questions than those we address here, but used the same task, and here we asked whether the model could offer an account of the complex interaction of interruption length and offset plotted in Figure 4. The lines in Figure 4 are the mean theoretical values from the model, averaged over individuals. Here we summarize how the model accounts for our data, and in the Appendix present the model formalisms.

The model accounts for our data in terms of differences in the operating principles of semantic and episodic memory. Semantic memory stores a stable long-term representation of the task sequence in which each step is associatively linked to its successor. Focusing on one step spreads activation to that step's immediate successor, much as focusing on a node in a semantic network spreads activation to neighboring nodes in the network. This spreading activation constitutes an implicit memory for the correct next step.

Spreading activation decays during interruptions, gradually returning the immediate successor to its base level of activation. If spreading activation were to decay completely, all steps of the task sequence would be at their base level of activation and therefore have an equal chance of being selected as the next step. Gradual decay toward this base level during interruptions causes a gradual linear increase in the sequence error rate at all offsets, as all incorrect steps grow more likely to intrude on the correct step. Spreading activation also reaches beyond the immediate successor, attenuating with each additional step it travels forward in the task sequence. This spread primes the step after the immediate successor, for example, causing it to intrude at an elevated rate, explaining why errors are more frequent at Offset $+1$ than at Offset $+2$.

Episodic memory stores a record of past performance that the system can sample for cues that can then serve to prime the next step in semantic memory. The contents of episodic memory decay, like spreading activation, but this decay continues without bound, to counteract buildup of proactive interference as the system continually encodes new episodic memories (Altmann, 2013; Altmann & Gray, 2002, 2008; Altmann & Schunn, 2012). This unbounded decay creates an activation ranking in which a memory for the most recent trial is the most active, a memory for the next most recent trial is the next most active, and so on. After an interruption, the system retrieves the most active item and uses it as a prime for the next step. Usually, the most active item will represent the preinterruption trial, which primes the correct successor. However, random fluctuation in activation levels (i.e., activation noise) means that sometimes the most active item will represent the prepre-interruption trial, which primes the preinterruption step and therefore leads to an error at Offset -1. Memories of older trials will have decayed enough that they are unlikely to be retrieved, so errors at Offsets -2 and -3 (and earlier, in the general case) are unlikely. Thus, retrieval errors from episodic memory mainly cause errors at Offset -1. In the Appendix we quantify the activation levels and retrieval probabilities associated with this analysis (see, in particular, Figure A1).

The effect of interruption length on retrieval errors from episodic memory, and thus sequence errors at Offset -1, depends on rehearsal. The model incorporates a representation of rehearsal in episodic memory, which participants often report using to keep placekeeping information active during interruptions (Altmann & Trafton, 2015; Trafton, Altmann, Brock, & Mintz, 2003). The model parameter governing rehearsal (E, described in the Appendix) is the time lag required to set up rehearsal at the start of the interruption. The longer this lag, the higher the error rate at Offset -1 on the postinterruption trial. The underlying mechanism is that during the lag, memory for the preinterruption trial decays relative to memory for the prepre-interruption trial, increasing the probability that a memory for the prepre-interruption trial intrudes when rehearsal begins. Once established, rehearsal maintains the rehearsed item in an active state for the balance of the interruption, while other items decay, ensuring that the rehearsed item primes the next step on the postinterruption trial. Thus, the error rate at Offset -1 on the postinterruption trial is affected by the duration of the lag to set up rehearsal but not by the duration of an interruption beyond this lag.

These decay and rehearsal mechanisms explain why Offset -1 responded differently to the interruption length manipulation than did other offsets (Figure 4). Specifically, the error rate at Offset −1 increased with interruption length until interruption length exceeded the lag to set up rehearsal, and then leveled off. The empirical trend across short, medium, and long interruptions is consistent with a theoretical leveling off because those lengths did not differ significantly (note the large error bars in Figure 4). These mechanisms predict that conditions that increase the lag to set up rehearsal, or that interfere with rehearsal after it is established, should increase postinterruption sequence errors mainly at Offset −1, a point we consider again below.

In sum, our data suggest that interruption length affected memory processes, and our model suggests that it affected more than one. There is an implicit memory for the next step to be performed, an explicit memory for the last step that was performed, and a rehearsal process that affects the explicit memory. As the implicit memory decays, all steps become more likely to intrude, affecting errors at all offsets. As the explicit memory decays, its predecessor becomes more likely to intrude, affecting error rates at Offset −1. The model is complex, but so is the empirical pattern, which does not resemble any single forgetting function. The model is also precisely formulated, as we show in the Appendix, and accounts for our data without modification from its original form (Altmann & Trafton, 2015). Finally, it makes a prediction, which is that blocking rehearsal during interruptions should affect sequence errors primarily involving repetition of the preinterruption step.

Practical Implications

Our data indicate that there are robust gradients to sequence errors in which the most likely errors after an interruption involve repeating or skipping a single step (errors at Offset −1 and Offset +1, respectively). This pattern holds across the full range of interruption lengths in our manipulation (Figure 4; see also Figure A2 in the Appendix, which emphasizes the gradients for both postinterruption and baseline trials). Thus, when a procedure involves a critical step that should not be skipped or repeated, such as administering a dose of medication, that step should be buffered against interruptions occurring immediately before or after.

Our results also suggest that interruptions influence performance primarily by affecting memory for the state of the task environment before the interruption. Sequence errors, which reflect loss of memory for past performance, increased substantially with interruption length. In contrast, nonsequence errors, which are independent of memory for past performance, were slightly but significantly less frequent on the postinterruption trial compared with baseline trials, and were unaffected by the interruption length manipulation. RT effects mirrored sequence error effects, linking the resumption lag often associated with interruptions (e.g., Grundgeiger, Sanderson, MacDougall, & Venkatesh, 2010) primarily to costs of memory retrieval. Accordingly, our results suggest that interruptions function as retention intervals and not so much as distractions that affect attention or other nonmemorial processes. This view suggests that interventions to reduce error and time costs of interruption should focus mainly on better support, or reduced need, for postinterruption memory retrievals (see also Edwards & Gronlund, 1998, and Oulasvirta & Saariluoma, 2006). This view also implies that people will sometimes forget the interrupted task altogether. In our procedure, the interrupted task was the only one available after an interruption, but this is not always true, and when there are other options people often choose them (Dodhia & Dismukes, 2009).

Our model suggests that the type of error most likely to be reduced by interventions that improve memory support will be incorrect repetitions of the preinterruption step (i.e., errors at Offset −1). The underlying mechanism is that errors in remembering preinterruption performance (as a prime for the next step) are most likely to involve confusions between preinterruption and prepre-interruption events. If the latter is mistaken for the former, the result will be a repetition of the preinterruption event. Maintaining an accurate memory for preinterruption performance, by whatever means, should minimize this particular confusion.

However, our model also suggests there are limits on what can be accomplished with improved memory support. The model assumes that knowledge representations required for task performance are primed through task performance, in that focusing on task-related elements (steps of the procedure, here) spreads activation to associated knowledge elements. The model also assumes that this spreading activation decays during interruptions, in a way that is not affected by memory strategies like rehearsal. If these assumptions are correct, meaning that the best way to activate task representations is through actual task performance, then interruption costs may be difficult to mitigate by any means other than restructuring the task to avoid interruptions.

At the same time, decay of spreading activation suggests that there may be at least a limited benefit of interruptions in reasoning or problem-solving tasks. If search through a space of possible solutions to a problem has reached a dead end, and this dead end is represented mentally in terms of a particular set of activated knowledge elements and intermediate products, the model predicts that an interruption could allow this activated context to decay and therefore allow search to take a more productive path afterward.

One specific activity that our results address is counting, a basic cognitive tool that people use in a wide variety of everyday situations. In the lab, counting errors have been studied by showing participants a sequence of events, such as asterisks appearing on the screen, and asking them to report the total count after the last event occurs (e.g., Carlson & Cassenti, 2004). A task that involves counting events or objects—cars on the road, repetitions of an exercise, people in a venue with a capacity limit—is procedural if one views the natural numbers as the task sequence and the operation of updating the count as the selection of the next step.

Our data suggest that when a counting task is interrupted, the type and range of errors in the total count at the end of the task will depend on the number and length of the interruptions during the task. A series of short interruptions should lead to a modest undercount, because the most likely error after each interruption is a single step backward in the natural numbers (an error at Offset −1). As interruption length increases, so should the probability of a wholesale loss of accuracy, as errors at all offsets become more frequent. Viewed as a task sequence, the natural numbers have a practically large and theoretically infinite number of steps, so the variance associated with a wholesale loss of accuracy after a long interruption should be much greater than in our task.

Limitations

One limitation of our task is that placekeeping is entirely memory based, whereas real-world task environments often offer explicit memory support, such as "to do" lists and checklists (see, e.g., Gawande, 2010). Such support does not negate a role for memory in placekeeping, if only because the acts of recording something on a list and referring to a list are themselves steps that can be skipped (see also Dismukes & Berman, 2010). Nonetheless, perceptual placekeeping clearly plays a role in many task environments, and measuring and modeling its effects is an important goal for future work.

Other limitations concern the task's timing parameters. For example, the baseline time to perform a step was just over 2 s (Figure 2), and although some procedural tasks unfold this quickly, many do not. Our model predicts that more time between steps should increase the temporal distinctiveness and thus the accuracy of memory for performed steps (Altmann & Hambrick, in press), thereby reducing errors at Offset −1. Thus, our empirical results may overestimate the rate of such errors generally. A second example is the interval between one interruption and the next, which was 10 to 15 s. Even in highly dynamic environments such as intensive care units, interruptions occur every couple of minutes (Spooner, Corley, Chaboyer, Hammond, & Fraser, 2015) rather than every few seconds. We would expect that interruptions that are less frequent, or that are engrossing in some way (Einstein, McDaniel, Williford, Pagan, & Dismukes, 2003), would be associated with less frequent use of active memory strategies for maintaining placekeeping information during interruptions. Our model predicts that less maintenance of placekeeping information during interruptions should increase errors at Offset −1, in which case our empirical results may also underestimate the rate of such errors generally. These conflicting estimates highlight both the need to understand differences between source and target tasks when generalizing results, and the role of a cognitive model in helping the analyst decide what those differences might mean.

One characteristic of our task that may or may not be a limitation is its abstract nature. Nothing in the materials, the task sequence, or the choice rules bears any resemblance to real-world task environments we know of. In contrast, other laboratory tasks are often designed with cover stories that add face validity, as in the simulated military planning task we have used in previous interruption research (e.g., Altmann & Trafton, 2007). However, we would argue that a cover story adds little actual external validity unless it preserves the complexity of authentic versions of the task, and therefore a role for training and knowledge. Cover stories also introduce the possibility that results are tied in some way to cover story details, and may add error variance if background knowledge about the cover story varies across individuals. Thus, we would argue that an abstract task has advantages over a task that looks more authentic than it really is.

In recent work, we have taken a differential approach to validating the UNRAVEL task, asking whether variability in UNRAVEL performance across individuals predicts variability in performance of other tasks. In one study (Hambrick & Altmann, 2015) we found that UNRAVEL performance predicts individual differences in general cognitive ability (Gf), which has predictive validity for a wide range of other tasks. The mechanistic connection between placekeeping and Gf, and possible predictive validity of UNRAVEL for specific workplace tasks, are issues to be explored in future work.

References

Altmann, E. M. (2011). Testing probability matching and episodic retrieval accounts of response repetition effects in task switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 935–951. http://dx.doi.org/10.1037/a0022931

Altmann, E. M. (2013). Fine-grain episodic memory processes in cognitive control. *Zeitschrift fur Psychologie, 221,* 23–32. http://dx.doi.org/10.1027/2151-2604/a000127

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13,* 27–33. http://dx.doi.org/10.1111/1467-9280.00405

Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review, 115,* 602–639. http://dx.doi.org/10.1037/0033-295X.115.3.602

Altmann, E. M., & Hambrick, D. Z. (in press). Practice increases procedural errors after task interruption. *Journal of Experimental Psychology: General*.

Altmann, E. M., & Schunn, C. D. (2012). Decay versus interference: A new look at an old interaction. *Psychological Science, 23,* 1435–1437. http://dx.doi.org/10.1177/0956797612446027

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26,* 39–83. http://dx.doi.org/10.1207/s15516709cog2601_2

Altmann, E. M., & Trafton, J. G. (2007). Timecourse of recovery from task interruption: Data and a model. *Psychonomic Bulletin & Review, 14,* 1079–1084. http://dx.doi.org/10.3758/BF03193094

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

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J. R., & Pirolli, P. (1984). Spread of activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 791–798. http://dx.doi.org/10.1037/0278-7393.10.4.791

Carlson, R. A., & Cassenti, D. N. (2004). Intentional control of event counting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1235–1251. http://dx.doi.org/10.1037/0278-7393.30.6.1235

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404–431. http://dx.doi.org/10.1037/0033-295X.97.3.404

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104,* 123–147. http://dx.doi.org/10.1037/0033-295X.104.1.123

Dismukes, R. K., & Berman, B. (2010). *Checklists and monitoring in the cockpit: Why crucial defenses sometimes fail*. (NASA/TM-2010–216396). Retrieved from http://human-factors.arc.nasa.gov/publications/NASA-TM-2010-216396.pdf

Dodhia, R. M., & Dismukes, R. K. (2009). Interruptions create prospective memory tasks. *Applied Cognitive Psychology, 23,* 73–89. http://dx.doi.org/10.1002/acp.1441

Edwards, M. B., & Gronlund, S. D. (1998). Task interruption and its effects on memory. *Memory, 6,* 665–687. http://dx.doi.org/10.1080/741943375

Einstein, G. O., McDaniel, M. A., Williford, C. L., Pagan, J. L., & Dismukes, R. K. (2003). Forgetting of intentions in demanding situations is rapid. *Journal of Experimental Psychology: Applied, 9,* 147–162. http://dx.doi.org/10.1037/1076-898X.9.3.147

Gawande, A. (2010). *The checklist manifesto: How to get things right*. New York, NY: Picador.

Grundgeiger, T., Sanderson, P., MacDougall, H. G., & Venkatesh, B. (2010). Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. *Journal of Experimental Psychology: Applied, 16,* 317–334. http://dx.doi.org/10.1037/a0021912

Grundgeiger, T., Sanderson, P. M., & Dismukes, R. K. (2014). Prospective memory in complex sociotechnical systems. *Zeitschrift fur Psychologie, 222,* 100–109. http://dx.doi.org/10.1027/2151-2604/a000171

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

Hodgetts, H. M., & Jones, D. M. (2006). Interruption of the Tower of London task: Support for a goal activation approach. *Journal of Experimental Psychology: General, 135,* 103–115. http://dx.doi.org/10.1037/0096-3445.135.1.103

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24,* 849–878. http://dx.doi.org/10.1017/S0140525X01000103

Katidioti, I., Borst, J. P., & Taatgen, N. A. (2014). What happens when we switch tasks: Pupil dilation in multitasking. *Journal of Experimental Psychology: Applied, 20,* 380–396. http://dx.doi.org/10.1037/xap0000031

Li, S. Y. W., Blandford, A., Cairns, P., & Young, R. M. (2008). The effect of interruptions on postcompletion and other procedural errors: An account based on the activation-based goal memory model. *Journal of Experimental Psychology: Applied, 14,* 314–328. http://dx.doi.org/10.1037/a0014397

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2004). Recovering from interruptions: Implications for driver distraction research. *Human Factors, 46,* 650–663. http://dx.doi.org/10.1518/hfes.46.4.650.56816

Neill, W. T. (1997). Episodic retrieval in negative priming and repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1291–1305. http://dx.doi.org/10.1037/0278-7393.23.6.1291

Oulasvirta, A., & Saariluoma, P. (2006). Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human-Computer Studies, 64,* 941–961. http://dx.doi.org/10.1016/j.ijhcs.2006.04.006

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

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of forgetting. *Psychological Review, 103,* 734–760. http://dx.doi.org/10.1037/0033-295X.103.4.734

Spooner, A. J., Corley, A., Chaboyer, W., Hammond, N. E., & Fraser, J. F. (2015). Measurement of the frequency and source of interruptions occurring during bedside nursing handover in the intensive care unit: An observational study. *Australian Critical Care, 28,* 19–23. http://dx.doi.org/10.1016/j.aucc.2014.04.002

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies, 58,* 583–603. http://dx.doi.org/10.1016/S1071-5819(03)00023-5

Trafton, J. G., Altmann, E. M., & Ratwani, R. (2011). A memory for goals model of sequence errors. *Cognitive Systems Research, 12,* 134–143. http://dx.doi.org/10.1016/j.cogsys.2010.07.010

Westbrook, J. I., Woods, A., Rob, M. I., Dunsmuir, W. T., & Day, R. O. (2010). Association of interruptions with an increased risk and severity of medication administration errors. *Archives of Internal Medicine, 170,* 683–690. http://dx.doi.org/10.1001/archinternmed.2010.65

Appendix

A Formal Cognitive Model of Placekeeping in the UNRAVEL Task

Here we describe the model that we fit to our data. The model was developed by Altmann and Trafton (2015) to address questions unrelated to those we address here. We applied it here to see if it could help explain the complex interaction of interruption length and offset evident in Figure 4. We describe the model in full but refer the reader to Altmann and Trafton (2015) for a more complete account of underlying assumptions.

The first section provides an overview of the basic theoretical constructs. The second section describes the model parameters. The third section presents the equations that specify activation levels and retrieval probabilities of memory codes and the associated probabilities of sequence errors. The fourth section describes the model fitting procedure and an assessment of model fit. The model, empirical data, and figure and table sources are posted online at msu.edu/~ema/interruptionlength/.

Model Overview

The model is based on two main theoretical ideas about the relationship between memory and performance. The first is that controlled performance on each trial of a procedure like ours is guided by a code in episodic memory that represents the processing context for that trial (Hommel, Musseler, Aschersleben, & Prinz, 2001; Neill, 1997). Because this code is represented in episodic memory, it persists for a time after performance of the trial is complete. The set of such control codes represents a history of recent performance that is available to help guide future performance (Altmann, 2011, 2013; Altmann & Gray, 2008; Trafton, Altmann, & Ratwani, 2011).

We identify each such control code as a *predecessor*, indexed according to the distance of the associated trial from the present. For example, $pred_1$ is from the most recently performed trial,

(*Appendix continues*)

$pred_2$ is from the trial before that, and so on. Predecessors decay over time, creating an activation ranking in which *$pred_1$* is the most active, *$pred_2$* the next most active, and so on. Decay of predecessors continues without bound, so that old predecessors effectively vanish from the system, making room for new ones (Altmann & Gray, 2002, 2008; Altmann & Schunn, 2012).

The second theoretical idea is that task knowledge is represented as an associative chain that functions much like any semantic network in terms of activation spreading between nodes. Each step (node) in the chain is associatively linked to its *successor*, which is indexed relative to whichever step currently occupies the system's focus of attention; *$succ_1$* is the immediate successor to the step in focus, *$succ_2$* is the next successor, and so on. The step in the focus of attention serves as an activation source that spreads activation forward through the chain. Spreading activation attenuates as it spreads, creating an activation ranking in which *$succ_1$* is the most active, *$succ_2$* is the next most active, and so on. Spreading activation also decays over time when the focus of attention shifts to another task during interruptions. Unlike predecessors, successors decay only to their base level of activation, which they have because they are stable elements of long-term knowledge.

Under these operating principles, the system can select the next step with high accuracy by performing two successive memory retrievals. The first retrieves the most active predecessor as a prime to focus on and the second retrieves the most active (primed) successor. The retrieved codes are usually *$pred_1$* and *$succ_1$*, respectively, but activation noise can transiently alter the activation ranking of predecessors or successors, causing a memory error that in turn causes a sequence error. For example, if *$pred_2$* happens to be transiently more active than *$pred_1$*, *$pred_2$* will be retrieved and then prime its *$succ_1$*, which is the step the system just performed; performing that step again would be a sequence error at Offset -1. For uniformity we assume that the same mechanisms govern placekeeping on the postinterruption trial and on baseline trials, and use data from both trial types to constrain the model.

The decay functions for predecessors and successors are both negatively accelerating, meaning that decay during interruptions increases interference. Specifically, during interruptions *$pred_1$* loses more activation than older predecessors and *$succ_1$* loses more activation than more distant successors, causing older predecessors and more distant successors to be more likely to intrude on *$pred_1$* or *$succ_1$*, respectively. However, the decay functions differ in that predecessors decay without bound whereas successors decay only to their base level of activation (see Figure A1, discussed in detail in the Model Equations section). Among predecessors, then, only *$pred_2$* has any real chance of being retrieved in place of *$pred_1$*, because older predecessors are too decayed to intrude. In contrast, any successor has some chance of being retrieved in place of a decayed *$succ_1$*, because all successors have a base level of activation associated with being stable elements of knowledge. Thus, decay of *$pred_1$* during an interruption affects mainly errors at Offset -1 on the postinterruption trial, whereas decay of *$succ_1$* during an interruption affects errors at all offsets on the postinterruption trial.

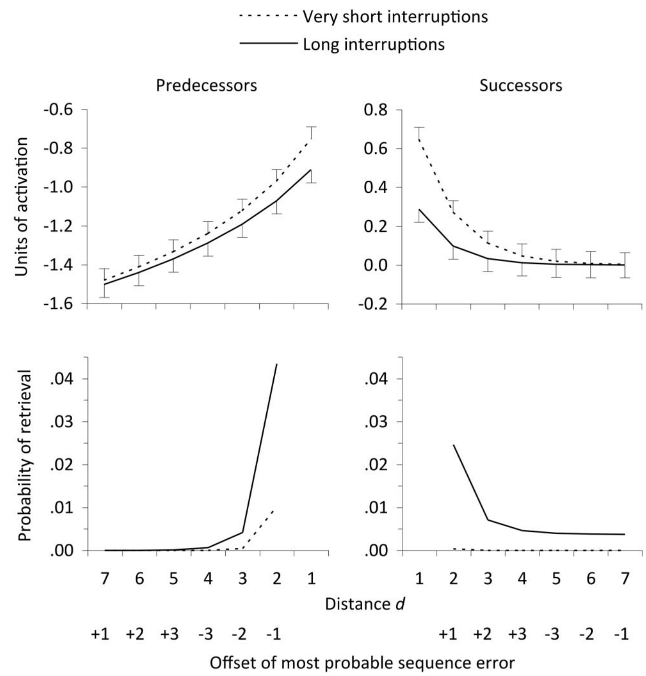


Figure A1. Activation functions (upper panels) and retrieval probabilities (lower panels) for the postinterruption trial, plotted using the mean parameter values from the very short and long interruptions conditions (see Table A1). Error bars are ± 1 standard deviation of activation noise. Upper left panel: Activation of predecessors (Equation A1). Upper right panel: Activation of successors (Equation A4). Lower left panel: Retrieval probability of predecessors (Equation A3) for $d > 1$. Lower right panel: Retrieval probability of successors (Equation A5) for $d > 1$.

A common strategy for maintaining placekeeping information in an active state during interruptions is rehearsal (Altmann & Trafton, 2015; Trafton et al., 2003), which we model by assuming that rehearsal targets *$pred_1$* (Altmann & Trafton, 2015). There is a lag to establish rehearsal, during which *$pred_1$* decays relative to *$pred_2$*, increasing the chances that *$pred_2$* interferes and ends up being the code that is rehearsed. Once established, rehearsal maintains the rehearsed code for the rest of the interruption, and the rehearsed code then primes its successor on the postinterruption trial. Thus, the lag to establish rehearsal at the start of the interruption—not directly the interruption length itself—affects the probability that *$pred_2$* is rehearsed instead of *$pred_1$* and thus the probability of a sequence error at Offset -1 on the postinterruption trial.

Model Parameters

The model comprises equations that characterize activation levels and retrieval probabilities for predecessors and successors. We fit these equations to each participant's data by estimating the values of five free parameters. We describe the parameters in this section and the equations in the next section.

(*Appendix continues*)

The first parameter, E, represents the lag to set up rehearsal at the start of the interruption. During this lag, $pred_1$ decays relative to $pred_2$, increasing the chances of an error at Offset −1 on the postinterruption trial. Accordingly, E is the *e*ffective interruption duration as far as errors at Offset −1 are concerned. E is one factor determining predecessor age in Equation A2, defined in the next section. We have no direct measure of E, so estimates of E are largely driven by the error rate at Offset −1 on the postinterruption trial.

The second parameter, s, governs noise in activation levels. Noise is operationalized as a random sample from a zero-mean logistic distribution taken independently for every memory element on every system cycle and added to that element's total activation. s is related by a transformation to the standard deviation of the distribution, so a larger s means more noise. s affects the mapping from activation levels to retrieval probabilities in Equations A3 and A5.

The third and fourth parameters, W_{post} and W_{base}, are two variants of the W parameter in Equation A4, which is the amount of activation spreading from the step in the focus of attention to its successors. W_{post} is for the postinterruption trial and absorbs decay of spreading activation during interruptions (Altmann & Trafton, 2015). W_{base} is for baseline trials.

The fifth parameter, g, is the proportion of spreading activation reaching a step that is passed on to that step's successor, again in Equation A4. Smaller values of g mean that less activation reaches distant successors.

Model Equations

Here we describe the model equations—first those for predecessors, then those for successors, and finally those that map retrieval probabilities to sequence error probabilities.

Predecessor activation and retrieval probability. To select the next step the system first retrieves a predecessor to use as a prime. The system retrieves the most active predecessor and simply assumes it has retrieved $pred_1$, which it usually has except when $pred_2$ has intruded.

Three equations jointly determine the retrieval probability of a predecessor. Equation A1 determines a predecessor's activation level from its age. Equation A2 determines a predecessor's age from various timing parameters. Equation A3 maps activation levels of the set of predecessors to the probability of retrieving any one of them.

The activation of $pred_d$—the control code that governed performance of the dth preceding trial—is given by

$$A(t_d) = -0.5\ln(t_d), \tag{A1}$$

where t_d is the age of $pred_d$ and 0.5 is the decay rate (Anderson & Lebiere, 1998). The upper left panel of Figure A1 shows that $A(t_d)$ decreases without bound as d increases (to the left).

The lower left panel of Figure A1 shows the corresponding retrieval probabilities (defined by Equation A3, below) for incorrect predecessors ($pred_{d>1}$). The retrieval probability after long versus very short interruptions differs substantially for $pred_2$, differs much less for $pred_3$, and differs not at all for older predecessors, which have no effective probability of intruding. Thus, as we wrote in the body, interruption length affects error rates at Offset −1 differently than it affects error rates at other offsets, because interruption length affects episodic memory, and most episodic memory errors involve retrieval of $pred_2$.

t_d is determined by timing parameters that include average response time for baseline trials R, the interruption duration I, and the effective interruption duration E:

$$t_d = \begin{cases} dR + E & \text{for Position 1} \\ dR & \text{for Position} \geq 2 \text{ and Position} \geq d \\ dR + I & \text{for Position} \geq 2 \text{ and Position} < d. \end{cases} \tag{A2}$$

The *position* factor that we introduce here is the serial position of the current trial after the interruption. d, again, is the distance of the predecessor from the present, measured in trials.

The first clause of Equation A2 applies to the postinterruption trial (Position 1) and determines t_d for all predecessors of that trial. For example, $pred_1$ at this point is from the preinterruption trial, so its age is $R + E$. $pred_2$ is older by a trial, so its age is $2R + E$.

The second and third clauses of Equation A2 apply to baseline trials (Position ≥ 2). The second clause determines t_d for all postinterruption predecessors (Position > d) and the preinterruption predecessor (Position = d). For example, for Position 2, $pred_1$ is from the postinterruption trial, and its age is simply R. $pred_2$ is from the preinterruption trial, and by assumption was rehearsed during the interruption (as $pred_1$, at that point). For tractability we assume that rehearsal exactly offsets decay during the interruption, so that for the rehearsed code the interruption in effect never happened. Thus, the age of $pred_2$ is now $2R$. The third clause determines t_d for predecessors preceding the preinterruption trial (Position < d). For example, for Position 2, $pred_3$ is from the prepre-interruption trial, and its age is $3R + I$, where $3R$ is the total response time for 3 consecutive trials and I is the interruption duration. By assumption, the code from the prepre-interruption trial was not rehearsed and therefore decayed for the full I.

The activation levels of all predecessors affect the probability of retrieving any one of them. The probability $u(d)$ of retrieving the dth predecessor is given by

$$u(d) = \frac{e^{A(t_d)/s}}{\sum_{i=1}^{D} e^{A(t_i)/s}}, \tag{A3}$$

where A is activation from Equation A1, $D = 7$ is the number of steps in the procedure, and $s = \sqrt{6}\sigma/\pi$, where σ is the standard deviation of activation noise (Anderson & Lebiere, 1998). This equation normalizes the activation of a candidate retrieval target to the total activation of all candidates, amplifying activation differences with exponentiation. The amplification is smaller with larger values of s.

(*Appendix continues*)

Successor activation and retrieval probability. After the system retrieves a predecessor, it focuses on that step to prime retrieval of that step's successor—though this priming also spreads to more distant successors. Two equations jointly determine the retrieval probability of a successor. Equation A4 specifies the initial amount of activation spreading from the source and the rate of attenuation as it spreads, and allows a representation of decay of spreading activation during interruptions. Equation A5, like Equation A3, maps activation levels to retrieval probabilities.

The activation spreading to $succ_d$ is given by

$$B(d) = Wg^{d-1}, \tag{A4}$$

where W is the initial amount of activation spreading from the source, g is the proportion of activation reaching one step that is passed to the next, and d is the number of steps from the source to $succ_d$, the dth most distant successor in the UNRAVEL sequence. The formalism is taken from Anderson and Pirolli (1984). As shown in the upper right panel of Figure A1, $B(d)$ approaches asymptote at 0 as d increases (to the right); 0 is the base level of activation of the task sequence.

The lower right panel of Figure A1 shows the corresponding retrieval probabilities (defined by Equation A5) for incorrect successors ($succ_{d\,>\,1}$). After long interruptions (solid line), all successors have a measurable probability of intruding on $succ_1$, because they maintained their base level activation as $succ_1$ decayed. This increase in the probability of all successors intruding accounts for the gradual linear increase in errors at all offsets in Figure 4.

To represent decay of spreading activation, we estimate W separately for the postinterruption trial (W_{post}) and for baseline trials (W_{base}), allowing W_{post} to absorb the effect of decay. The underlying theoretical assumption is that the rate of decay at a step is proportional to the amount of activation that has spread to that step (Altmann & Trafton, 2015).

The probability $v(d)$ of retrieving the dth successor is

$$v(d) = \frac{e^{B(d)/s}}{\sum_{i=1}^{D} e^{B(i)/s}}, \tag{A5}$$

where B is from Equation A4 and D and s are as in Equation A3.

Predicted sequence error rates. The retrieved predecessor and successor determine which step is selected for performance. For example, $pred_1$ and $succ_1$ generate the correct next step—but so do $pred_2$ and $succ_2$, as do all other pairs that represent canceling retrieval errors. The joint probability of any path other than $pred_1$ and $succ_1$ is very small, so earlier we treated any retrieval other than $pred_1$ and $succ_1$ as an error, but for completeness we include all alternative paths in the model. Thus, the probability of selecting the correct next step is a sum of probabilities,

$$p_{correct} = \sum_{d=1}^{D} u(d)v(d), \tag{A6}$$

where here as earlier $D = 7$ so that each step nominally has a chance to influence performance.

Similarly, the most probable path to an error at Offset -1 is through $pred_2$ and $succ_1$, as we have discussed, but another path is through $pred_3$ and $succ_2$, and yet another is through $pred_1$ and $succ_7$. Generalizing this logic, the probability of an error at Offset $-n$ is

$$p_{-n} = \sum_{d=1}^{D} u[f(n,d,D)]v(d) \tag{A7}$$

and the probability of an error at Offset $+n$ is

$$p_{+n} = \sum_{d=1}^{D} u(d)v[f(n,d,D)], \tag{A8}$$

where $f(n, d, D) = 1 + [(d - 1 + n) \bmod D]$. The values for p_{-3}, p_{-2}, p_{-1}, p_{+1}, p_{+2}, and p_{+3} are the theoretical error proportions at offsets -3, -2, -1, $+1$, $+2$, and $+3$, respectively.

Procedures for Fitting and Evaluating the Model

We fit the model using maximum likelihood estimation based on the binomial distribution,

$$Likelihood = \binom{n}{k} p^k (1-p)^{n-k}, \tag{A9}$$

where, for a given cell of the experimental design, n is the actual number of trials, k is the actual number of sequence errors, and p is the error probability for that cell predicted by the model. We used the Microsoft Excel Solver routine to estimate values for E, s, W_{post}, W_{base}, and g that maximized the sum of log likelihoods across the 36 cells formed by crossing the six levels of offset (-3, -2, -1, $+1$, $+2$, $+3$) with six levels of position (1 through 6), where position is the serial position of a trial after an interruption (Position 1 being the postinterruption trial and Positions 2 through 6 being baseline trials). We substitute the position factor for the trial type factor used in the body because the model is constrained by the empirical error rate at each of these levels of position.

We fit the model to the data from each participant individually, generating theoretical values to code alongside that participant's empirical values. The means of the theoretical values across participants are the lines in Figure 4 and also in Figure A2, which shows a more comprehensive comparison of theoretical and empirical values that includes baseline trials (averaged over Positions 2 through 6) as well as the postinterruption trial. Table A1 gives mean parameter values for each interruption length. The mean parameter values for the very short and long conditions were used to generate the functions in Figure A1.

(*Appendix continues*)

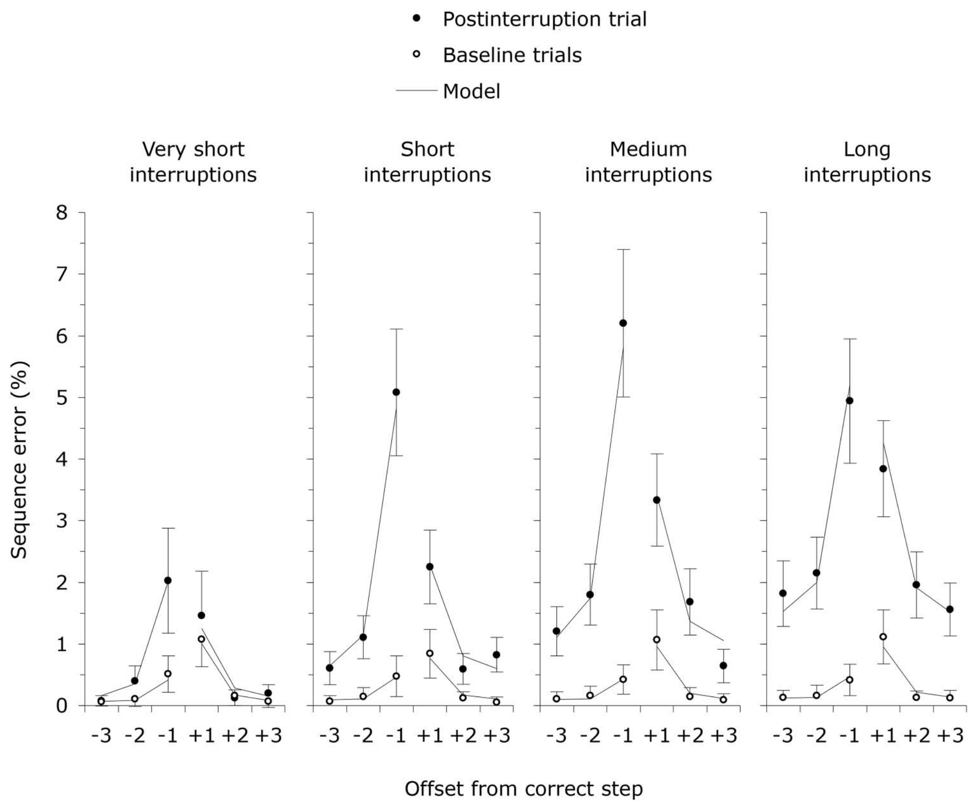


Figure A2. Sequence errors on the postinterruption trial and on baseline trials. Markers are empirical values, error bars are 95% confidence intervals, and lines are model values.

To evaluate the model's goodness of fit we used an inferential test developed by Altmann and Trafton (2015). The test augments the empirical ANOVA design with an additional within-participants factor called *fit*, with levels *empirical* (participant data) and *theoretical* (participant model values). An interaction of a contrast in the empirical design with the fit factor indicates that the model cannot account for that empirical effect. Altmann and Trafton (2015) found the test powerful enough to detect fairly subtle model-data misfits.

(*Appendix continues*)

Table A1
Mean Parameter Values for the Model for Each Interruption Length Condition

	Free parameters					Bound parameters	
Interruption length	E	s	W_{post}	W_{base}	g	R	I
Very short	2.02	0.048	0.65	0.56	0.42	2.76	2.76
Short	4.19	0.052	0.39	0.50	0.32	2.74	13.12
Medium	4.47	0.052	0.32	0.47	0.34	2.63	22.02
Long	3.86	0.052	0.29	0.46	0.34	2.63	31.91

Note. Values for free parameters are estimated through model fitting: E is in seconds, s is related by a transformation to units of activation (see text), W_{post} and W_{base} are in units of activation, and g is dimensionless. Values of bound parameters are bound by performance data: R is response time averaged over Positions 2 through 6, and I is the interruption duration. R and I are means of untrimmed participant means, in seconds.

Table A2 shows the test applied to the present data. Columns 1 through 7 show results of a 4 (interruption length) × 6 (offset) × 6 (position) ANOVA on the empirical data. Columns 8 and 9 show results of the goodness-of-fit test. Each F ratio in Column 9 is formed from the effect term in Column 8 and the error term in Column 5. The effect term represents the deviation of theoretical means from empirical means for that empirical contrast. The error term is simply the estimate of the empirical error variance; the model adds no variance of its own, so pooling estimates across levels of the fit factor only attenuates the error variance, the more so the better the fit. The F ratio is significant if the deviation of theoretical means from empirical means is large relative to the error variance in the data. No Fs in Column 9 were greater than 1, so there is no basis to reject the model on grounds that it failed to account for systematic variance attributable to experimental factors.

Table A2
Analysis of Variance for Sequence Errors (Columns 1–7) and the Corresponding Model Goodness-of-Fit Test (Columns 8–9)

1	2	3	4	5	6	7	8	9
							Contrast × Fit	
Contrast	MS_{effect}	MS_{error}	df_{effect}	df_{error}	F	p	MS_{effect}	F
Interruption length (L)	89.45	8.34	3	396	10.73	.000	0.17	0.02
Position (P)	1012.73	3.81	5	1,980	265.60	.000	1.71	0.45
Offset (O)	595.00	3.14	5	1,980	189.59	.000	2.14	0.68
L × P	80.70	3.81	15	1,980	21.16	.000	0.72	0.19
L × O	6.79	3.14	15	1,980	2.16	.006	0.62	0.20
P × O	111.29	2.28	25	9,900	48.77	.000	1.20	0.53
L × P × O	6.85	2.28	75	9,900	3.00	.000	1.02	0.45

Received June 11, 2016
Revision received December 1, 2016
Accepted December 3, 2016 ■