

## Practice Increases Procedural Errors After Task Interruption

Altmann, E. M. & Hambrick, D. Z.

### *Supplemental Materials*

Here we present two analyses of the Altmann and Trafton (2015) model (see also Altmann, Trafton, & Hambrick, in press) that produced the theoretical values in Figure 4. First, we map the formalisms that explain our data to those of a model of temporal distinctiveness (Brown, Neath, & Chater, 2007). The Altmann and Trafton model is grounded in decay theory, which is generally viewed as a competitor to distinctiveness theory (e.g., Brown et al., 2007; Crowder, 1976; Lewandowsky, Oberauer, & Brown, 2009), but, as we show here, a fleshed-out decay theory is formally identical to distinctiveness theory for the special case in which the retrieval target is the most recent item.

Second, we evaluate the model's goodness-of-fit using an inferential test developed by Altmann and Trafton (2015). The main result of this analysis is that the model provides a zero-parameter fit of the session effect on post-interruption sequence errors. That is, the full magnitude of the increase in post-interruption sequence errors between sessions is accounted for by the decrease in baseline response time (RT) between sessions, to within a level of tolerance determined by the error variance in the data.

The model and our data are posted online at: [msu.edu/~ema/practice/](http://msu.edu/~ema/practice/)

### **Mapping the Model to Temporal Distinctiveness**

The model comprises closed-form equations that map the ages of memory codes to their activation levels, activation levels to retrieval probabilities, and retrieval probabilities to sequence error probabilities. Here we focus on the equations that link the decrease in baseline RT between sessions to the increase in post-interruption sequence errors between sessions.

The activation of  $pred_d$ , the memory for the  $d$ th “predecessor” (i.e., preceding trial), is

$$A_d = -0.5 \ln(t_d), \quad (S1)$$

where 0.5 is the decay rate and  $t_d$  is the age of  $pred_d$ . The age  $t_d$ , computed via Equation A2 of Altmann and Trafton (2015) with  $L = 0$ , factors in the average RT on baseline trials, which corresponds to the interpresentation interval in the ratio rule.

The probability  $u_d$  of retrieving  $pred_d$  is

$$u_d = \frac{e^{\frac{A_d}{s}}}{\sum_{i=1}^D e^{\frac{A_i}{s}}}, \quad (S2)$$

where  $A$  is the activation level from Equation S1,  $s$  governs activation noise, and  $D$  is the total number of steps in the procedure (7, for UNRAVEL). Equations S1 and S2 are taken from the ACT-R cognitive theory (Anderson & Lebiere, 1998), with Equation S1 representing decay and Equation S2 representing one form of interference.

Equations S1 and S2 are isomorphic to Equations 1 and 5 of the SIMPLE model of temporal distinctiveness (Brown et al., 2007), for the case in which the retrieval target is the most recent item. That case applies here, because the target is  $pred_1$ , the memory for the most recently performed trial.

The mapping is as follows. Equation S2 represents absolute activation values, but SIMPLE emphasizes differences on a psychological dimension. Transformed to represent differences in activation between target  $pred_1$  and distractors  $pred_d$ , Equation S2 becomes,

$$u_d = \frac{e^{\frac{A_d - A_1}{s}}}{\sum_{i=1}^D e^{\frac{A_i - A_1}{s}}}. \quad (S3)$$

In place of activation values  $A$ , SIMPLE specifies memory locations  $M$ . Both quantities are log-transformations of time that give more weight to a difference in age of neighboring items the

more recent the items are. The difference is that  $A$  is a negative log transform (activation decreases with time) modified by a decay rate, whereas  $M$  is a positive log transform (distance of a memory location from the present increases with time). Accordingly, Equation S3 can be transformed to represent memory locations with the substitution  $A = -0.5M$ , producing

$$u_d = \frac{\eta_{d,1}}{\sum_{i=1}^D \eta_{i,1}}, \quad (\text{S4})$$

where

$$\eta_{i,1} = e^{\frac{-0.5}{s}(M_i - M_1)}. \quad (\text{S5})$$

The retrieval target  $pred_1$  occupies the most recent memory location,  $M_1$ . Because all other items are located further away than  $M_1$ , the difference  $(M_i - M_1)$  in Equation S5 is always nonnegative and is therefore equivalent to  $|M_i - M_1|$ . With this substitution, Equation S5 is equivalent to SIMPLE Equation 1 and Equation S4 is equivalent to SIMPLE Equation 5, for the case in which the retrieval target is the most recent item. This mapping suggests that decay and temporal distinctiveness have more in common conceptually than has been acknowledged to date (e.g., Lewandowsky et al., 2009).

### Testing Model Fit

As in Altmann and Trafton (2015) and Altmann et al. (in press), we fit the model to each participant's data using maximum likelihood estimation. Here we estimated 5 free parameters to maximize the sum of log likelihoods across the 72 cells of a 2 (session) x 6 (offset) x 6 (position) design. Table S1 shows mean estimated parameter values, along with means for the bound parameters  $R$  (empirical RT) and  $I$  (empirical interruption duration). Altmann and Trafton describe the functional role of each parameter in detail.

Of relevance here is that the model offers a zero-parameter account of the increase in post-interruption sequence errors between sessions. Two model parameters ( $R$  and  $I$ ) vary between sessions, but both are bound by empirical data, and we claim that together they fully account for the increase in sequence errors. This claim rests on an inferential goodness-of-fit test showing that the model leaves no experimental variance in the sequence error data unexplained. The test, developed by Altmann and Trafton (2015), involves augmenting the analysis of variance (ANOVA) for the experimental design with an additional within-participants factor called *fit*, with levels *empirical*, meaning the sequence errors generated by each participant, and *theoretical*, meaning the corresponding model values. The test produces one  $F$ -ratio for each interaction of the fit factor with a contrast (main effect or interaction) in the experimental design. A significant  $F$  indicates that model-data deviations for that particular contrast differed from 0, implying that the model was not able to accommodate that particular experimental effect.

Table S2 shows the test results. Columns 1...7 are the results of a 2 (session) x 6 (position) x 6 (offset) ANOVA on empirical sequence error proportions, and columns 8 and 9 constitute the goodness-of-fit test. Each  $F$ -ratio in column 9 is formed from the effect term in column 8 and the empirical error term in column 5. A significant  $F$  in column 9 would mean that the difference between empirical and theoretical means is large relative to the error term. No  $F$  in column 9 was greater than 1. In particular,  $F < 1$  for all contrasts involving session, licensing the conclusion that the model leaves no systematic variance due to session unexplained. This full accounting of the experimental results, mediated by the memory mechanisms represented in the model, supports a causal interpretation linking the session effect on RT to the session effect on post-interruption sequence errors.

That said, as with any ANOVA, this test is sensitive to heterogeneity of variance—which we have in our data, with the error term for the session effect being greater on the post-interruption trial than on baseline trials (compare the error bars for Position 1 versus Positions 2...6 in the top panel of Figure 3). The goodness-of-fit test reported above was therefore more powerful for the post-interruption trial than it would have been had error variance not been pooled across all levels of position. By the same token, the test was less powerful for baseline trials than it would have been without pooling—and, indeed, the test failed to detect the fact that the model does not account for the session effect on baseline sequence errors (compare the bottom panels of Figure 4).

When we tested goodness-of-fit on baseline trials alone, excluding the post-interruption trial and its associated error variance, the Fit x Session interaction was, in fact, marginally significant,  $F(1, 205) = 3.02, p = .084$ , indicating that the model did not track the session effect. We leave it to future work to diagnose the exact problem with the model's account of baseline performance. For present purposes, the fact that the test could detect the misfit for baseline performance with an appropriate error term builds confidence that the null results in column 9 of Table S2 reflect the model's accuracy as an account of post-interruption performance, which is the focus of our study.

## 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.
- Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (in press). Effects of interruption length on procedural errors. *Journal of Experimental Psychology: Applied*.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Erlbaum.
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114, 539-576.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). Response to Altmann: Adaptive forgetting by decay or removal of STM contents? *Trends in Cognitive Sciences*, 13, 280-281.

*Table S1: Model Parameter Values and Descriptions.*

Name	$M$	$SD$	Description
Free parameters:			
$E$	3.61	1.91	Seconds from interruption onset to rehearsal onset
$s$	0.051	0.014	Transformed $SD$ of activation noise
$W_{post}$	0.27	0.10	Spreading activation at source, post-interruption trial
$W_{base}$	0.42	0.30	Spreading activation at source, baseline trials
$g$	0.27	0.10	Proportion of activation spreading to next step
Bound parameters:			
$R_1$	2.65	0.51	RT (seconds) on Positions 2...6, Session 1
$R_2$	2.14	0.47	RT (seconds) on Positions 2...6, Session 2
$I_1$	22.31	6.43	Interruption duration (seconds), Session 1
$I_2$	20.66	5.98	Interruption duration (seconds), Session 2

*Note.*  $M$ ,  $SD$  - mean and standard deviation across participants. RT - response time.

*Table S2: Results of a 2 (Session) x 6 (Position) x 6 (Offset) 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 x Fit	
Contrast	$df_{effect}$	$MS_{effect}$	$df_{error}$	$MS_{error}$	$F$	$p$	$MS_{effect}$	$F$
Session (S)	1	1.66E-07	205	4.05E-04	0.00	.984	2.67E-04	0.66
Position (P)	5	1.57E-01	1025	5.48E-04	286.02	.000	3.40E-05	0.06
Offset (O)	5	4.01E-02	1025	4.28E-04	93.76	.000	3.01E-04	0.70
S x P	5	1.45E-03	1025	2.34E-04	6.20	.000	1.54E-04	0.66
S x O	5	4.67E-04	1025	1.99E-04	2.35	.039	5.10E-05	0.26
P x O	25	1.23E-02	5125	2.06E-04	59.90	.000	1.69E-04	0.82
S x P x O	25	5.04E-04	5125	1.85E-04	2.72	.000	8.00E-05	0.43