A Correction to the SIMPLE Model of Free Recall

James P. Pooley
University of California, Irvine
Michael D. Lee
University of California, Irvine

Abstract

The SIMPLE (Scale-Invariant Memory, Perception, and LEarning) model developed by Brown, Neath, and Chater (2007) has provided a formal instantiation of the idea that scale-invariance is an important organizing principle across numerous cognitive domains, and has made an influential recent contribution to the cognitive science literature dealing with modeling human memory. In this note, we correct a previously unreported mistake in the specification of the SIMPLE model, within the context of free recall data. We show that the error matters not only in theory, but also in practice, by re-applying the corrected SIMPLE model to the classic data set of Murdock (1962).

Introduction

Brown et al. (2007) introduced the SIMPLE (Scale-Invariant Memory, Perception, and LEarning) model as a formal instantiation of the idea that scale-invariance is an important organizing principle across numerous cognitive domains (Kello et al., 2010). The model has been applied to empirical phenomena observed throughout the domains of perception, learning, and memory. Despite this potentially wide applicability, our focus in this note is the verbal free recall paradigm, where the model has received attention recently (e.g., Farrell, 2010; Laming, 2010; Lewandowsky & Farrell, 2011).

In this note, we correct a previously unreported mistake in the specification of the SIMPLE model, within the context of free recall data. We show that the error matters not only in theory, but also in practice, by re-applying the corrected SIMPLE model to the classic data set of Murdock (1962).

Address correspondence to: James P. Pooley, Department of Cognitive Sciences, University of California, Irvine, Irvine, CA, 92697-5100. Electronic Mail: jpooley@uci.edu

The SIMPLE Model

In this section, we provide a brief review of the psychological assumptions and formal structure of the SIMPLE model that pertain to free recall. Nothing we present in this section is a novel development, and we include this section only to make this note self contained. We also take the opportunity to standardize and correct the mathematical notation used to define the model. For more information on SIMPLE, the reader should consult Brown et al. (2007).

Representation of Items

The SIMPLE model relies on the geometric assumption common to many psychological models that stimulus items can be represented as points in a multidimensional "psychological space" (e.g., Nosofsky, 1992; Shepard, 1980, 1987). Although it is possible to construct such a space using multidimensional scaling algorithms (e.g., Borg & Groenen, 2005; Shepard, 1974), it is typically assumed that a unidimensional space, with the single dimension corresponding to a logarithmically compressed representation of time, is adequate in the context of free recall. Assuming such a unidimensional representation of items in memory, the psychological distance between the ith and jth item representations (or memory locations) is given by the absolute difference

$$d_{ij} = |x_i - x_j| \,, \tag{1}$$

where $x_i = \ln(T_i)$ is the representation (or location) of the ith item, presented at time T_i relative to the start of the period of recall, in psychological space.

Item Similarity

Given these psychological distances, the pairwise similarity between the ith and jth item representations (or memory locations) is given by

$$\eta_{ij} = \exp(-c d_{ij}) \,, \tag{2}$$

where the parameter $c \in [0, \infty)$ determines the distinctiveness (temporal or otherwise) of the item representations or memory locations, with larger values corresponding to greater distinctiveness. Shepard (1987) provides experimental and theoretical motivation for this similarity function.

Retrieval Probability

A core psychological assumption of the SIMPLE model is that items are retrieved from memory on the basis of their discriminability to a given retrieval cue, where discriminability is assumed to be given by a similarity choice rule of the form (e.g., Luce, 1963)

$$p_{ij} = \frac{\eta_{ij}}{\sum_{k=1}^{n} \eta_{ik}} . \tag{3}$$

In all of the applications of the SIMPLE model of which we are aware, it has been assumed that the study list provides the relevant contextual cues. In other words, the set of retrievable items is taken to be the set of n study list items.

To account for the omission errors that typically occur during free recall, Brown et al. (2007, p. 545) argue that "[i]ntuition suggests that items that are difficult to discriminate should be most likely to be omitted. Any discriminabilities that fall below some threshold value will lead to omissions, and any that fall above the threshold will lead to overt recalls. Assuming some noise in activations or thresholds, this will have the overall effect of increasing recall probabilities that are already high and reducing recall probabilities for items whose recall probabilities are already low."

This assumption is implemented in the SIMPLE model by adopting a logistic transfer function of the form

$$p^*_{ij} = \frac{1}{1 + \exp\left[-s(p_{ij} - t)\right]}, \tag{4}$$

where the parameters $s \in [0, \infty)$ and $t \in [0, 1]$ are intended to measure, respectively, the scale (or "noise") and threshold of their hypothesized "threshold mechanism."

Response Process

In free recall—as opposed to, say, serial recall—an item can be recalled correctly in any output position. This means that Equation 4 is insufficient for calculating the probability that any given item will ultimately be recalled. Brown et al. (2007) appear to conceptualize free recall as a sequential process much like serial recall. In their formal definition of the SIMPLE model, the probability of recalling the ith of the n words on the study list is given as

$$\theta_i = \min\left(1, \sum_{j=1}^{n} p^*_{ij}\right). \tag{5}$$

This corresponds to aggregating by summation the retrieval probability of a word over a set of potential retrieval attempts, as made clear in the worked computational example provided in the appendix of Brown et al. (2007)

A Correction to the SIMPLE Model

As with any model of a psychological process, the SIMPLE model relies on drastic simplifications of complex psychological processes, and these assumptions are open to criticism. Our goal in this note is not to provide such a critique. Equations 1–4 accurately formalize the assumptions of those parts of the model. This note, instead, targets the final Equation 5, which incorrectly formalizes the probability of recall.

Brown et al. (2007, p. 545) justify Equation 5 as follows: "The mechanism described [by Equation 4] has the consequence that zero or small values of discriminability can lead to higher predicted recall probabilities. In some applications, such as serial recall, this can occasionally result in a predicted sum of recall probabilities slightly greater than 1.0 for a given output position; *rather than adopt a more complex thresholding mechanism we simply normalize or cap recall probabilities to 1.0 in such circumstances* [emphasis added]."

This is not correct. Imagine a tourist planning a weekend visit to Warwick castle, and hoping that at least one of the two days will be free of rain. If the probability of any given day at the castle being free of rain is 70%, then what is the probability the tourist will get their rain free day? Clearly, it is not $0.7 + 0.7 = 1.4$, nor is it 1.4 thresholded to 1.0. Instead, the probability of at least one rain free day is naturally calculated by first finding the probability that both days are rainy $(1 - 0.7) \times (1 - 0.7) = .09$, and taking the complement $1.0 - .09 = 0.91$.

The first approach, based on direct summation and thresholding, is incorrect, but is exactly what is used by Equation 5 in the SIMPLE model. The correct approach is given by elementary probability theory. Applied to the problem of determining the recall probability θ_i for the ith item, based on n retrieval attempts, the jth of which has probability p^*_{ij} of success, is given by

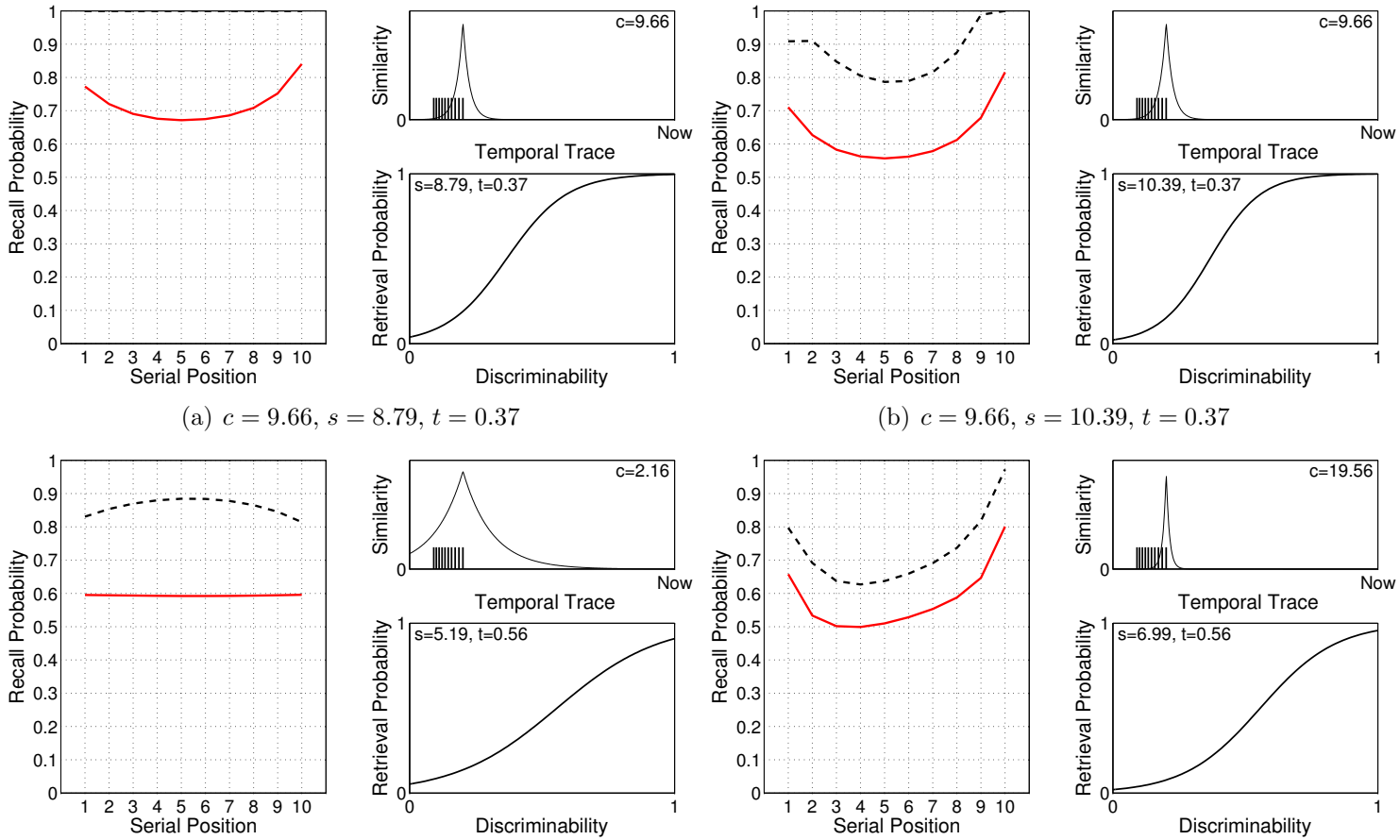
$$\theta_i = 1 - \prod_{j=1}^{n} \left(1 - p^*_{ij}\right). \tag{6}$$

In all extant applications of SIMPLE of which we are aware (e.g., Lewandowsky & Farrell, 2011; Pooley, Lee, & Shankle, 2011), Equation 5 has been used to apply the model to free recall data, with the additive form of the retrieval probabilities typically justified by the assumed sequential structure of the retrieval process.[1]

Behavior With and Without Correction

Figure 1 presents a series of explorations of the SIMPLE model's behavior in its original form, and with the correction. Each of the four panels correspond to a different combination of model parameters, for the same sequence of 10 memory items. Within each panel, the main plot on the left shows the serial position curve generated by the original model (broken black line) and the corrected model (solid red line) at those parameters. The smaller upper plot shows the logarithmically compressed memory item representations as vertical bars, and overlays the generalization gradient corresponding to the c parameter, as given by Equation 2. The small lower plot shows the logistic transfer function defined by the s and t parameters.

[1] We note, however, that presented with the same problem in his own model, Laming (2010, p. 126) correctly observes that the "[c]alculation of serial position curves for total recall proceeds on the basis of failure to retrieve a given word at any point during recall (i.e., the complement of the quantities conventionally modeled)." Interestingly, Laming discusses the SIMPLE model, but does not discuss its need for correction.



(a) $c = 9.66$, $s = 8.79$, $t = 0.37$

(b) $c = 9.66$, $s = 10.39$, $t = 0.37$

(c) $c = 2.16$, $s = 5.19$, $t = 0.56$

(d) $c = 19.56$, $s = 6.99$, $t = 0.56$

Figure 1. Behavior of the original and corrected SIMPLE model for free recall under four different parameter combinations. In each case the large left plot shows the serial position curve for the original (solid red line) and corrected (solid red line) versions, and the smaller plots show the memory representation with generalization gradient, and the transfer function.

It is clear from the four examples presented in Figure 1 that the serial position curve generated by the original SIMPLE model is, for a range of parameter combinations, very different from the curve generated by the corrected model. It is also clear that the original SIMPLE model generates counter-intuitive serial position curves. In panel (a) the original model predicts that all words will be recalled perfectly, while the corrected model predicts a more plausible pattern of recall, showing the expected primacy and recency effects. The problem with the original model comes from the threshold in Equation 5. The behavior of the original SIMPLE model in panels (b) and (c) is less obviously affected by thresholding, but still produces unusual serial position curves, whereas the corrected model produces what appear to be sensible predictions. The example in panel (d) shows that, even when both the original and corrected versions of SIMPLE produce serial position curves with standard qualitative primacy and recency components, the quantitative details of the curves differ. Thus, it is not that the original version of SIMPLE only produces some degenerate predictions at extreme values of parameters.

Based on the examples in Figure 1, and examining a wide range of other parameterizations, we believe that the original SIMPLE often produces very different serial position curves from the corrected version. Some of these are degenerate or implausible, but many are simply quantitatively different. Accordingly, the original version of SIMPLE makes erroneous predictions about data that may often be difficult to detect.

Applying the Model to Data

A key application of the SIMPLE model in Brown et al. (2007) involves the seminal data collected by (Murdock, 1962), which serves as a benchmark for the serial position effect in free recall. These data give the proportion of words correctly recalled, averaged across participants, for lists of 10, 15, and 20 words presented at a rate of 1 s per word, and lists of 20, 30 and 40 words presented at a rate of 2 s per word.

Brown et al. (2007) fit the SIMPLE model to data using using least squares optimization, finding parameter values that best match recall probabilities from the model to recall proportions observed in data. An alternative, and superior, Bayesian approach for fitting the model is provided by Shiffrin, Lee, Kim, and Wagenmakers (2008). The issue of the correctness of model specification that this note addresses is orthogonal to the issue of how statistical inference is done. We use the Bayesian approach because it is principled and coherent. Thus, following Shiffrin et al. (2008), we use non-informative prior distributions for the parameters: λ, $\sigma \sim \text{Uniform}\,(0, 100)$ and $\tau \sim \text{Beta}\,(1, 1)$.

Bayesian inference for SIMPLE was accomplished using JAGS (Plummer, 2003), free software which utilizes a variety of Markov chain Monte Carlo (MCMC) sampling algorithms (e.g., Gamerman & Lopes, 2006) to obtain the joint posterior distribution of the unknown quantities of interest. The JAGS implementation of SIMPLE, showing

the one line that needs to be changed to switch between the original and corrected version of the model, is shown below.

```
# SIMPLE Model
model{
   # Observed and predicted data
   for (k in 1:dsets){
      for (i in 1:listlength[k]){
         y[i,k] ~ dbin(theta[i,k],tr[k])
         predy[i,k] ~ dbin(theta[i,k],tr[k])
         predpc[i,k] <- predy[i,k]/tr[k]
      }
   }
   # Similarities, Discriminabilities, and Response Probabilities
   for (k in 1:dsets){
      for (i in 1:listlength[k]){

         for (j in 1:listlength[k]){
            # Similarities
            sim[i,j,k] <- exp(-c[k]*abs(log(m[i,k])-log(m[j,k])))
            # Discriminabilities
            disc[i,j,k] <- sim[i,j,k]/sum(sim[i,1:listlength[k],k])
            # Response Probabilities
            resp[i,j,k] <- 1/(1+exp(-s[k]*(disc[i,j,k]-t[k])));
         }
         # Free Recall Overall Response Probability
         theta[i,k] <- 1-prod(1-resp[i,1:listlength[k],k]) # Corrected Version
         # theta[i,k] <- min(1,sum(resp[i,1:listlength[k],k])) # Incorrect Version
      }
   }
   # Priors
   for (k in 1:dsets){
      c[k] ~ dunif(0,100)
      s[k] ~ dunif(0,100)
      t[k] ~ dbeta(1,1)
   }
}
```

Modeling Results

Model Checking

Evaluating the adequacy of a psychological model is a fundamental step in the psychological research process. Although numerous considerations motivate the evaluation of a psychological model (Shiffrin et al., 2008, pp. 1248–1249), all evaluation criteria require that the model achieve some basic level of descriptive adequacy. This is especially true in the case of simple models such as SIMPLE, where the primary purpose is to provide economical but principled accounts of data. Toward this end, we examine the posterior predictive distributions of the recall data. Posterior predictive distributions provide an intuitive and principled approach to assessing the descriptive

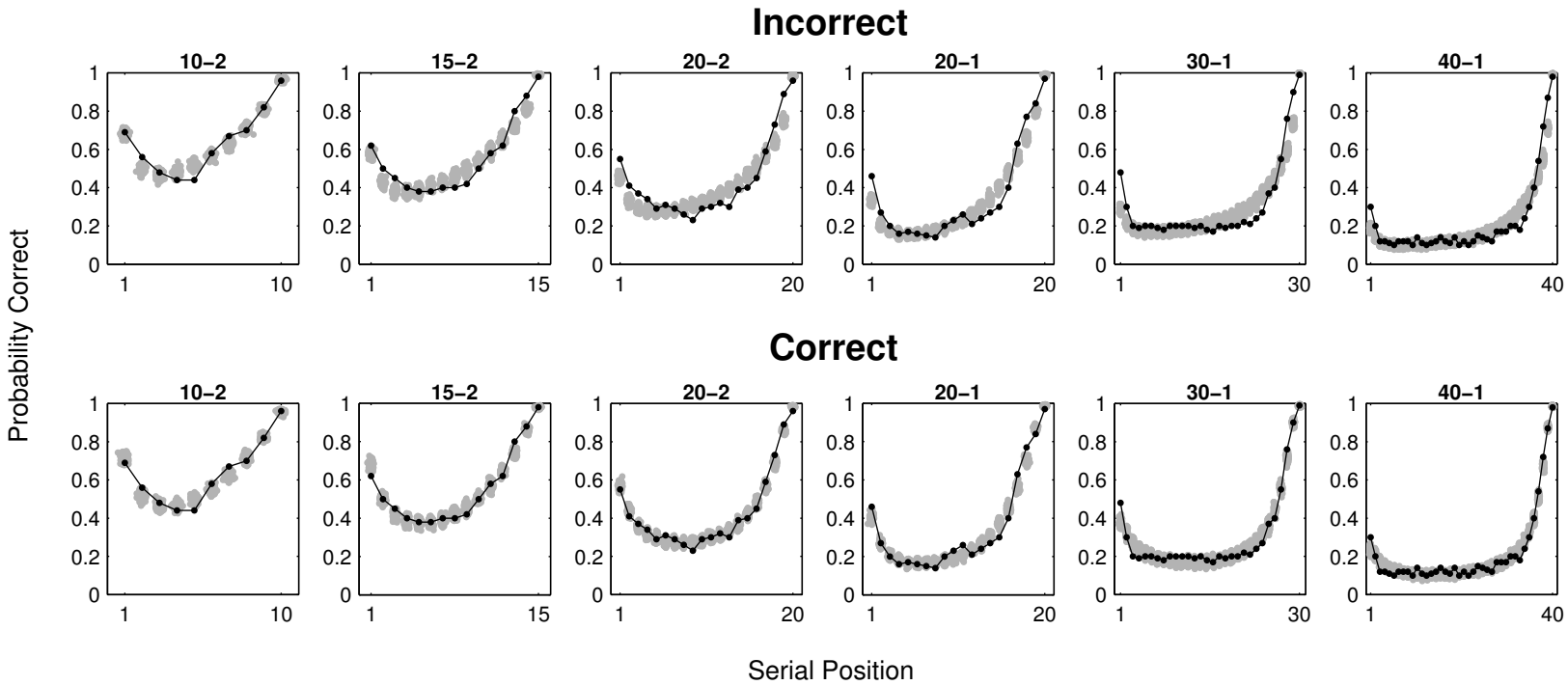


Figure 2. Posterior predictive transfer functions for the correct (left panel) and incorrect (right panel) implementations of SIMPLE, for each condition of the Murdock (1962) data.

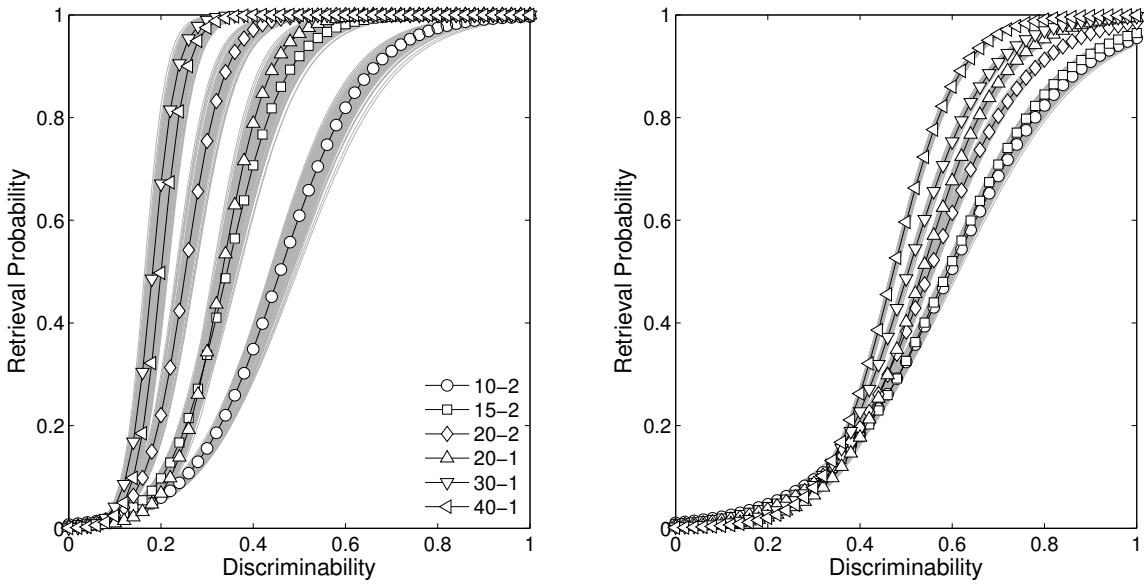


Figure 3. Posterior predictive transfer functions for the correct (left panel) and incorrect (right panel) implementations of SIMPLE, for each condition of the Murdock (1962) data.

adequacy of a Bayesian model (e.g., Gelman, Carlin, Stern, & Rubin, 2004, Chap. 6). Posterior predictions for some unobserved quantity correspond to numerical values generated by the model, using parameter values sampled from its posterior distribution, and naturally take into account uncertainty in these predictions.

Serial Position Curves

Figure 4 displays the posterior predictive distributions for the two implementations of SIMPLE. Clearly, both fit the data reasonably well, but the fit is much tighter for the correct implementation of the model.

Transfer Functions

Figure 3 displays the posterior predictive transfer functions for the correct (left panel) and incorrect (right panel) implementations of SIMPLE, for each condition of the Murdock (1962) data. Clearly, (i) there is more change over the conditions in the correct implementation of the model, (ii) the correct implementation of the model drops to much lower thresholds, as if people are willing to take anything more distinctive than some low threshold in the harder tasks (e.g., those in which the list length is reasonably large), and (iii) the orderings of the functions are sometimes different, and sometimes to a rather large extent (e.g., the 20-1 versus 20-2 conditions).

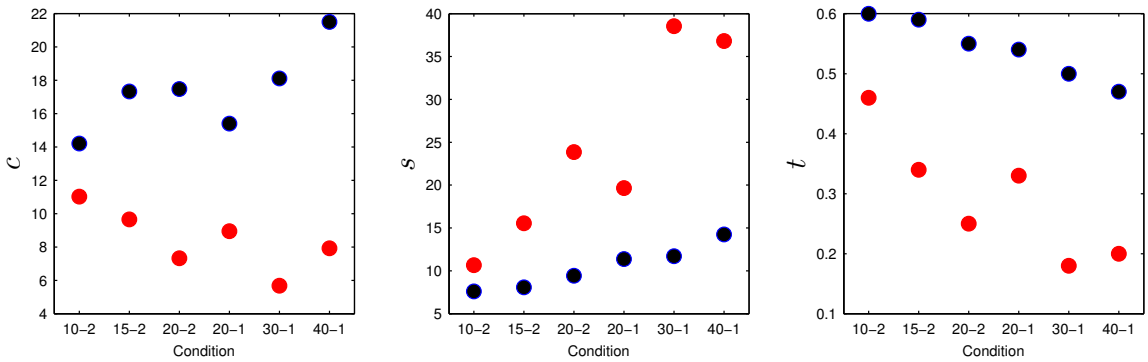


Figure 4. Posterior means for the parameters of SIMPLE for using the incorrect Equation 5 (blue, darker) and correct Equation 6 (red, lighter) for each condition of the Murdock (1962) data set.

Posterior Means

Since the posterior distributions for the parameters for each condition of the Murdock (1962) data do not overlap, it is reasonable to work with the posterior means rather than the full posterior distributions. Figure 4 displays these posterior means for the parameters of SIMPLE for each condition of the Murdock (1962) data set. Clearly, there are both qualitative and quantitative differences between the two implementations, and these differences are reflected in the better fits of the correct implementation obtained above.

Discussion

In this note we have corrected a previously unreported error in the implementation of SIMPLE as it apples to free recall data. We have also shown that, when correctly implemented, SIMPLE fits a classic free recall data set better and produces qualitative and quantitative changes in the parameter estimates obtained for the model. This raises a number of questions to be explored in future research.

Interpreting the Parameters

While not directly related to the correction reported in this note, it seems worthwhile to mention that Brown et al. (2007) never give psychological interpretations of the parameters estimates of SIMPLE in the free recall setting. Although they seem to motivate these parameters in psychological terms (e.g., "thresholds"), the parameter values shown in Figure 4 (e.g., s especially, but also t) often seem implausible. Although the situation is a bit cleaner for the correct implementation, questions still remain as to how these parameters (and, indeed, SIMPLE itself) should be interpreted (e.g., as a data-fitting model or a more process-oriented model).

Why Does This Matter?

Many current theoretical accounts of free recall give at least a cursory discussion of the principles of SIMPLE, even if they do not directly apply the model. It is assumed that these accounts have the incorrect implementation of the model in mind since Equation 5 is not called into question. The incorrect implementation of SIMPLE using Equation 5 also makes an appearance in the recent cognitive modeling textbook by Lewandowsky and Farrell (2011), where it is used to help explain the principles behind cognitive modeling. Given the use of the mode, then, and regardless of whether a researcher likes or dislikes the model, it seems worthwhile to point out that a key feature of this model is in error, which this note has done.

Acknowledgements

We thank Michael Kahana for making the Murdock (1962) data publically available.

References

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed.). New York: Springer.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*, 539–576.

Farrell, S. (2010). Dissociating conditional recency in immediate and delayed free recall: A challenge for unitary models of recency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 324–347.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis* (Second ed.). Boca Raton, FL: Chapman & Hall/CRC.

Kello, C. T., Brown, G. D. A., Cancho, R. F. i, Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., et al. (2010). Scaling laws in cognitive sciences. *Trends in Cognitive Sciences*, *14*, 223–232.

Laming, D. (2010). Serial position curves in free recall. *Psychological Review*, *117*, 93–133.

Lewandowsky, S., & Farrell, S. (2011). *Computational Modeling in Cognition: Principles and Practice*. Thousand Oaks, CA: Sage Publications.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 1, pp. 103–189). New York: Wiley.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, *43*, 25–53.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing.* Vienna, Austria.

Pooley, J., Lee, M., & Shankle, W. R. (2011). Modeling multitrial free recall with unknown rehearsal times. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, *39*, 373–422.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*, 390–398.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Shiffrin, R. M., Lee, M. D., Kim, W.-J., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*(8), 1248–1284.