15 The SIMPLE model of memory

15.1 The SIMPLE model

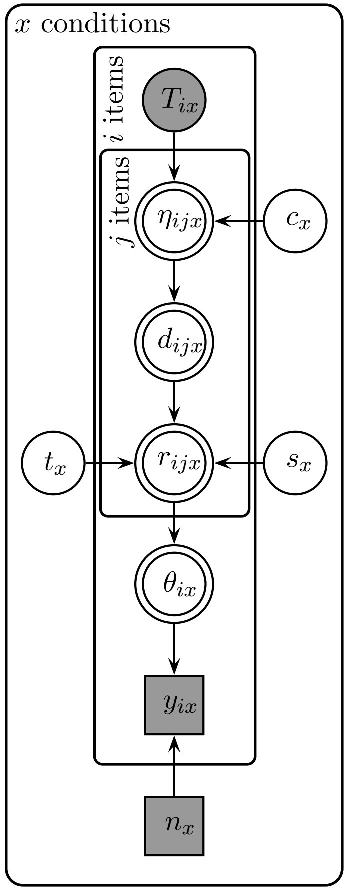
Brown, Neath, and Chater (2007) proposed the SIMPLE (Scale-Invariant Memory, Perception, and LEarning) model, which, among various applications, has been applied to the basic memory phenomenon of free recall. In this application, the SIMPLE model assumes memories are encoded by the time they were presented, but that the representations are logarithmically compressed, so that more temporally distant memories are more similar. It also assumes that distinctiveness plays a central role in performance on memory tasks, and that interference rather than decay is responsible for forgetting. Perhaps most importantly, the SIMPLE model assumes that the same memory processes operate at all time scales, unlike theories and models that assume different mechanisms for short-term and long-term memory.

The first application considered by Brown et al. (2007) involves seminal immediate free recall data reported by Murdock (1962). The data give the proportion of words correctly recalled averaged across participants, for lists of 10, 15, and 20 words presented at a rate of 2 seconds per word, and lists of 20, 30, and 40 words presented at a rate of 1 second per word.

Brown et al. (2007) make some reasonable assumptions about undocumented aspects of the task (e.g., the mean time of recall from the end-of-list presentation), to set the time T_i between learning and retrieval of the ith item. With these times established, the application of the SIMPLE model to the free recall data involves five stages, which are clearly described in Brown et al. (2007, Appendix).

First, the ith presented item, associated with time T_i, is represented in memory using logarithmic compression, given by $M_i = \log T_i$. Secondly, the similarity between each pair of items is calculated as $\eta_{ij} = \exp\left(-c\left|M_i - M_j\right|\right)$, where c is a parameter measuring the "distinctiveness" of memory. Thirdly, the discriminability of each pair of items is calculated as $d_{ij} = \eta_{ij} / \sum_k \eta_{ik}$. Fourthly, the retrieval probability of each pair of items is calculated as $r_{ij} = 1/\left(1 + \exp\left(-s\left(d_{ij} - t\right)\right)\right)$, where t is a threshold parameter and s is a threshold noise parameter. Finally, for free recall, the probability that the ith item in the presented sequence will be recalled is calculated as $\theta_i = \min\left(1, \sum_k r_{ik}\right)$.

These stages are implemented by the graphical model in Figure 15.1. The graphical model has variables corresponding to the observed times between learning and retrieval, T_i, and the observed number of correct responses y_{ix} for the ith item. The



$$
\begin{aligned}
c_x &\sim \text{Uniform}(0, 100)\\
s_x &\sim \text{Uniform}(0, 100)\\
t_x &\sim \text{Uniform}(0, 1)\\
\eta_{ijx} &\leftarrow \exp\left(-c_x \left|\log T_{ix} - \log T_{jx}\right|\right)\\
d_{ijx} &\leftarrow \eta_{ijx} / \textstyle\sum_k \eta_{ikx}\\
r_{ijx} &\leftarrow 1/\left(1 + \exp\left(-s_x \left(d_{ijx} - t_x\right)\right)\right)\\
\theta_{ix} &\leftarrow \min\left(1, \textstyle\sum_k r_{ikx}\right)\\
y_{ix} &\sim \text{Binomial}(\theta_{ix}, n_x)
\end{aligned}
$$

Fig. 15.1 The SIMPLE model of free recall.

available data are aggregated across subjects, and the different word lists completed by subjects, and so n_x counts the total number of trials for the xth condition, which is the same for all words in that condition. The similarity η_{ijx}, discriminability d_{ijx}, retrieval r_{ijx}, and free recall probability θ_{ix} variables are deterministic. They implement the SIMPLE model in terms of its three parameters, linking the temporal representation of the items to the behavioral data, which is the accuracy of recalling the items.

In Figure 15.1 the times, responses, and free recall probabilities apply per item, and so are enclosed in a plate replicating over items. The similarity, discriminability, and retrieval measures apply to pairs of variables, and so involve an additional plate also replicating over items. We follow Brown et al. (2007) by fitting the c, t, and s parameters independently for each condition. This means the entire graphical model is also enclosed in a plate replicating over the 6 conditions in the Murdock (1962) data.

Box 15.1 Adhockery

"Faced with a new problem, a classical statistician is free to invent new estimators, confidence intervals or hypothesis tests . . . In contrast, there is a unique Bayesian solution to any problem. That is the posterior distribution, which expresses the investigator's knowledge about θ after observing x. The Bayesian statistician's task is to identify the posterior distribution as accurately as possible, which usually entails identifying the prior distribution and the likelihood and then applying Bayes' theorem. There is no room for adhockery in Bayesian statistics." (O'Hagan & Forster, 2004, p. 19)

The script `SIMPLE_1.txt` implements the graphical model in WinBUGS. Note that the posterior predictive is calculated in some detail, leading up to `pcpred`, which is the posterior predicted proportion of correct recalls:

```
# SIMPLE Model
model{
  # Observed and Predicted Data
  for (x in 1:dsets){
    for (i in 1:listlength[x]){
      y[i,x] ~ dbin(theta[i,x],n[x])
      predy[i,x] ~ dbin(theta[i,x],n[x])
      predpc[i,x] <- predy[i,x]/n[x]
    }
  }
  # Similarities, Discriminabilities, and Response Probabilities
  for (x in 1:dsets){
    for (i in 1:listlength[x]){
      for (j in 1:listlength[x]){
        # Similarities
        sim[i,j,x] <- exp(-c[x]*abs(log(m[i,x])-log(m[j,x])))
        # Discriminabilities
        disc[i,j,x] <- sim[i,j,x]/sum(sim[i,1:listlength[x],x])
        # Response Probabilities
        resp[i,j,x] <- 1/(1+exp(-s[x]*(disc[i,j,x]-t[x])))
      }
      # Free Recall Overall Response Probability
      theta[i,x] <- min(1,sum(resp[i,1:listlength[x],x]))
    }
  }
  # Priors
  for (x in 1:dsets){
    c[x] ~ dunif(0,100)
    s[x] ~ dunif(0,100)
    t[x] ~ dbeta(1,1)
  }
}
```

The code `Simple_1.m` or `Simple_1.R` applies the model to the Murdock (1962) data. It involves some initial steps to get the data organized before passing it to

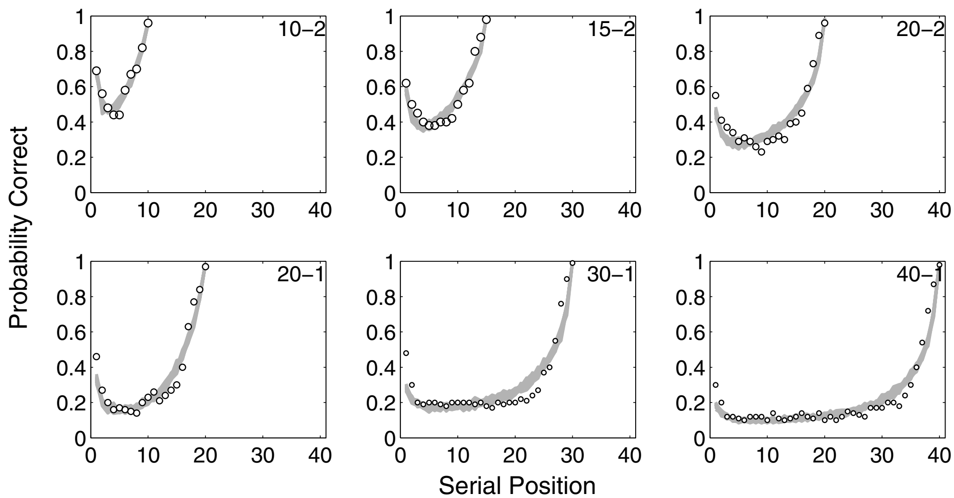


Fig. 15.2 Posterior prediction of the SIMPLE model for the six conditions of the Murdock (1962) immediate free recall data. The circles show the data, and the gray lines show the posterior predictive distribution.

WinBUGS. It also produces an analysis of the posterior predictions and the joint posterior parameter space.

The posterior predictive analysis for the six data sets is shown in Figure 15.2. The solid lines show the probability that the item in each serial position was correctly recalled. A total of 20 samples from the posterior predictive are shown for each serial position as gray points, making a gray area that spans the range in which the model expects the data to lie.

The posterior analysis is shown in Figure 15.3, which shows the joint posterior parameter distribution as a three-dimensional plot, with 20 posterior samples for each condition shown by different markers. The solid black dots projected onto the planes represent samples from the pairwise joint distributions of each possible combination of parameters, marginalized over the other parameter in each case. Finally, the marginal distributions for each parameter are shown along the three axes.

Figure 15.3 provides information about the distinctiveness, threshold, and threshold noise parameters, including information about variability and co-variation of the parameters across experimental conditions. This additional information is important to understanding how parameters should be interpreted, and for suggesting further model development. For example, the lack of overlap of the three-dimensional points for the six conditions suggests that there are important differences in model parameters for different item list lengths and presentation rates. In particular, it seems unlikely that an alternative approach to fitting the six conditions using a single discriminability level and threshold function will be adequate.

Box 15.2 Correcting the SIMPLE Model

Lee and Pooley (2013) point out that the part of the SIMPLE model that generates free recall probabilities is incorrect. The aim of this part of the model is to calculate the probability that an item will be recalled, and in free recall that is achieved if the word is recalled in any position in a recall sequence. Thus, all of the individual probabilities of recalling an item must be combined. The original formulation of the SIMPLE model does this additively, using $\theta_i = \min\left(1, \sum_x r_{ix}\right)$, with the thresholding at 1 guaranteeing that a probability is produced.

But, as Lee and Pooley (2013) point out, if the probability of an event is 0.7, then the probability that event will occur at least once on two independent trials is not $0.7 + 0.7 = 1.4$, nor is it 1.4 thresholded to 1.0. The correct probability is naturally calculated by first finding the probability that the event occurs on neither trial as $(1 - 0.7) \times (1 - 0.7) = 0.09$, and taking the complement $1.0 - 0.09 = 0.91$ for an item. Thus, Lee and Pooley (2013) argue for $\theta_i = 1 - \prod_x (1 - r_{ix})$. Implementing this correction in WinBUGS requires only replacing `theta[i,x] <- min(1,sum(resp[i,1:listlength[x],x]))` with `theta[i,x] <- 1-prod(1-resp[i,1:listlength[x],x])`.

Another intuition, this time coming from the two-dimensional joint posteriors, is that there is a trade-off between the threshold and threshold noise parameters, since their joint distributions (shown by the solid black dots in the bottom plane) show a high level of negative correlation for all of the conditions. This means that the data in each condition are consistent with relatively high thresholds and relatively low levels of threshold noise, or with relatively low thresholds and relatively high levels of threshold noise. This is probably not ideal, since generally parameters are more easily interpreted and theoretically compelling if they operate independently of each other. In this way, the information in the joint parameter posterior suggests an area in which the model might need further development or refinement.

As a final example of the information in the joint posterior, note that the marginal distributions for the threshold parameter shown in Figure 15.3 seem to show a systematic relationship with item list length. In particular, the threshold decreases as the item list length increases from 10 to 40, with overlap between the two conditions with the most similar lengths (i.e., the 10–2 and 15–2 conditions, and the 20–2 and 20–1 conditions). This type of systematic relationship suggests that, rather than treating the threshold as a free parameter, it can be modeled in terms of the known item list length.

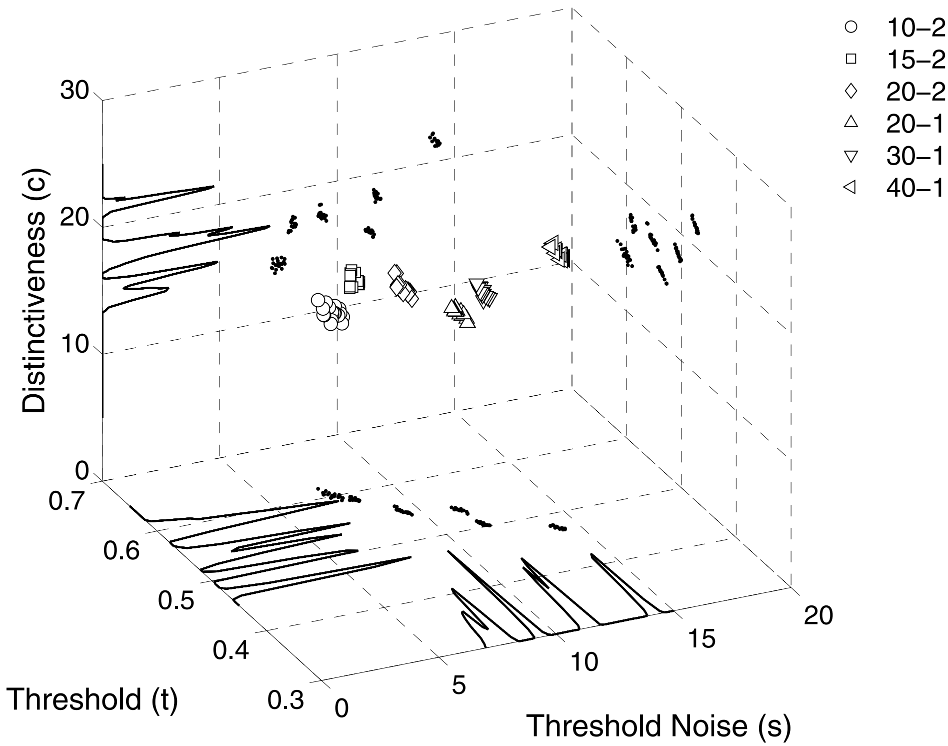


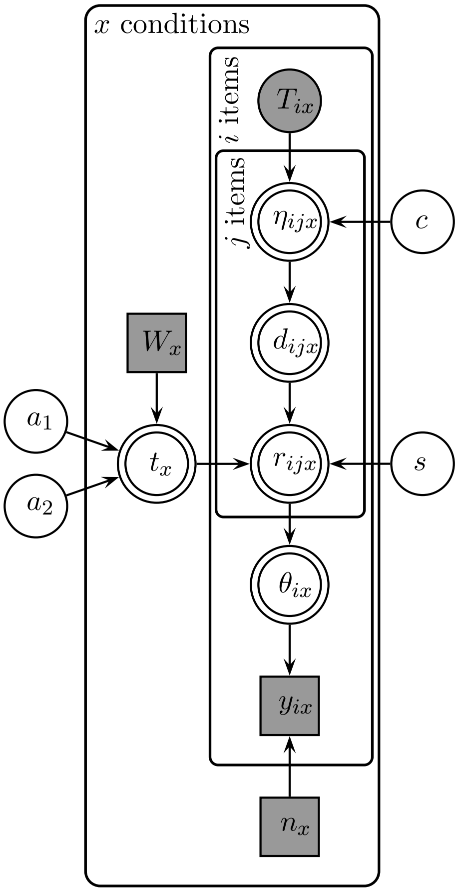
Fig. 15.3 Joint posterior parameter space for the SIMPLE model for the six conditions of the Murdock (1962) immediate free recall data.

Exercise

Exercise 15.1.1 Modify the graphical model so that the same parameter values are used to account for all of the data sets. You will also need to modify the Matlab or R code that produces the graphs.

15.2 A hierarchical extension of SIMPLE

We now consider how the possibility of a relationship between list length and thresholds can be implemented in a hierarchical extension to the SIMPLE model. The extended model is shown in Figure 15.4. There are two important changes. First, the distinctiveness c and threshold noise s parameters are now assumed to have the same value for all experimental conditions. In Figure 15.4, they are outside



$$
\begin{aligned}
c &\sim \text{Uniform}(0, 100)\\
s &\sim \text{Uniform}(0, 100)\\
a_1 &\sim \text{Uniform}(-1, 0)\\
a_2 &\sim \text{Uniform}(0, 1)\\
t_x &\leftarrow a_1 W_x + a_2\\
\eta_{ijx} &\leftarrow \exp\left(-c_x \left|\log T_{ix} - \log T_{jx}\right|\right)\\
d_{ijx} &\leftarrow \eta_{ijx} / \textstyle\sum_k \eta_{ikx}\\
r_{ijx} &\leftarrow 1/\left(1 + \exp\left(-s_x \left(d_{ijx} - t_x\right)\right)\right)\\
\theta_{ix} &\leftarrow \min\left(1, \textstyle\sum_k r_{ikx}\right)\\
y_{ix} &\sim \text{Binomial}(\theta_{ix}, n_x)
\end{aligned}
$$

Fig. 15.4 Graphical model implementing a hierarchical extension to the SIMPLE model of memory.

the plate that replicates over the experimental conditions, and they are no longer indexed by x.[1]

The second change corresponds to the way the thresholds t_x are determined. Rather than being assumed to be independent, these thresholds now depend on the item list length, denoted W_x for the xth condition. The dependence is modeled as a linear function $t_x = a_1 W_x + a_2$, parameterized by the coefficients a_1 and a_2. Consistent with the intuitions gained from Figure 15.3, we make the assumption that the linear relationship expresses a decrease in threshold as item list length increases, by using the prior $a_1 \sim \text{Uniform}(-1, 0)$.

[1] This is probably not a theoretically realistic assumption—indeed, as we pointed out, the joint posterior in Figure 15.3 argues against it—but it is a simple assumption that makes it easy to focus on the hierarchical extension.

The goal of the hierarchical extensions is to move away from thinking of parameters as psychological variables that vary independently for every possible recall task. Rather, we now conceive of the parameters as psychological variables that themselves now need explanation, and attempt to model how they change in terms of more general parameters.

This approach not only forces theorizing and modeling to tackle new basic questions about how recall processes work, but also facilitates evaluation of the prediction and generalization capabilities of the basic model (Ahn, Busemeyer, Wagenmakers, & Stout, 2008). By making the threshold parameter depend on characteristics of the task—in this case, the number of words in the list—in a systematic ways, and by treating the other parameters as invariant, the hierarchical extension automatically allows the SIMPLE model to make predictions about other tasks.

The script `SIMPLE_2.txt` implements the graphical model in WinBUGS:

```
# Hierarchical SIMPLE Model
model{
  # Observed data
  for (x in 1:dsets){
    for (i in 1:listlength[x]){
      y[i,x] ~ dbin(theta[i,x],n[x])
    }
  }
  # Similarities, Discriminabilities, and Response Probabilities
  for (x in 1:gsets){
    t[x] <- max(0,min(1,a[1]*w[x]+a[2]))
    for (i in 1:listlength[x]){
      for (j in 1:listlength[x]){
        # Similarities
        sim[i,j,x] <- exp(-c*abs(log(m[i,x])-log(m[j,x])))
        # Discriminabilities
        disc[i,j,x] <- sim[i,j,x]/sum(sim[i,1:listlength[x],x])
        # Response Probabilities
        resp[i,j,x] <- 1/(1+exp(-s*(disc[i,j,x]-t[x])))
      }
      # Free Recall Overall Response Probability
      theta[i,x] <- min(1,sum(resp[i,1:listlength[x],x]))
    }
  }
  # Priors
  c ~ dunif(0,100)
  s ~ dunif(0,100)
  a[1] ~ dunif(-1,0)
  a[2] ~ dunif(0,1)
  # Predicted data
  for (x in 1:gsets){
    for (i in 1:listlength[x]){
      predy[i,x] ~ dbin(theta[i,x],n[x])
      predpc[i,x] <- predy[i,x]/n[x]
    }
  }
}
```

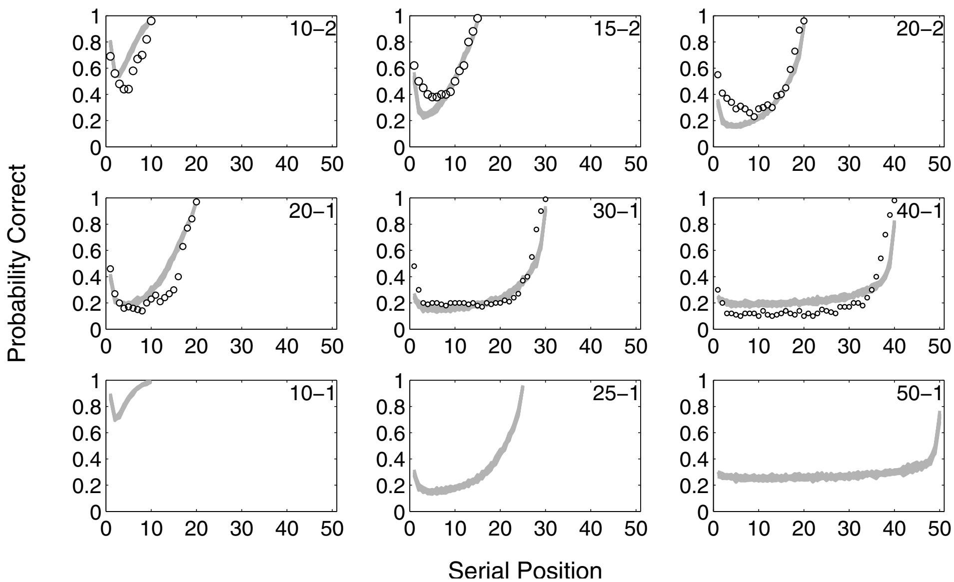


Fig. 15.5 Posterior prediction of the hierarchical extension of the SIMPLE model for the six conditions of the Murdock (1962) immediate free recall data, and in generalizing to three new conditions. The circles show the data, and the gray lines show the posterior predictive distribution.

The code `Simple_2.m` or `Simple_2.R` applies the hierarchical model to the Murdock (1962) conditions, and also to three other possible conditions, for which data are not available. These generalization conditions all involve presentation rates of 1 s per item, but with 10, 25, and 50 items, corresponding to both interpolations and extrapolations relative to the collected data.

The posterior predictive performance is shown in Figure 15.5. The top two rows show the Murdock (1962) conditions, while the bottom row shows the predictions the model makes about the generalization conditions. The model does not fit the data very well in the Murdock (1962) conditions, but our focus is on the possibility of the generalization predictions. Here, the hierarchical extension to the model allows it to predict serial recall curves for experimental conditions for which data are not available.

The posterior analysis in Figure 15.6 shows inferences about the threshold noise, distinctiveness, and threshold parameters. For the first two of these, the inferences take the form of single posterior distributions. For the threshold parameter, however, the posterior inference is now about its functional relationship to item list length. The posterior distribution for this function is represented in the right panel of Figure 15.6 by showing 50 posterior samples at each possible length $W = 1, \ldots, 50$. These posterior samples are found by taking joint posterior samples (a_1, a_2) and finding $t = a_1 W + a_2$ for all values of W.

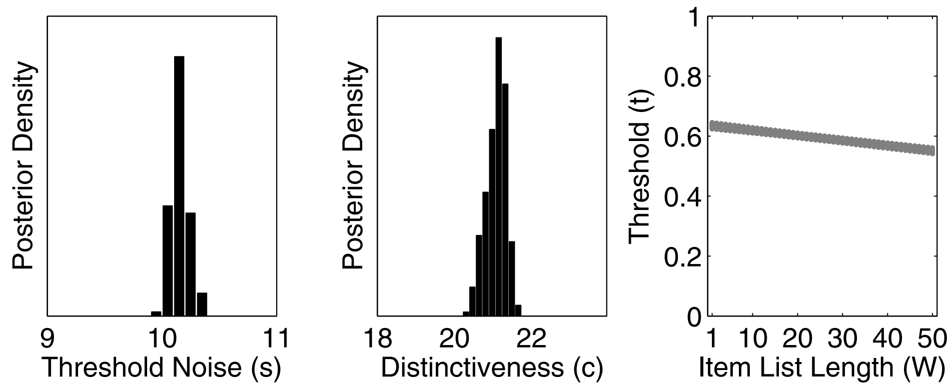


Fig. 15.6 Posteriors for the SIMPLE model parameters in its hierarchically extended form.

Exercise

Exercise 15.2.1 Why are empirical tests of generalization potentially more powerful or compelling evaluations of a model than fitting to existing data?