Modeling Multitrial Free Recall with Unknown Rehearsal Times

James P. Pooley (jpooley@uci.edu)
Michael D. Lee (mdlee@uci.edu)
Department of Cognitive Sciences, University of California, Irvine
Irvine, CA, 92697-5100

William R. Shankle (rshankle@mccare.com)
Medical Care Corporation, 19782 MacArthur Boulevard
Irvine, CA 92612

Abstract

Quantitative models of human memory often rely on assumed latent memory processes, such as patterns of rehearsal of the words on a study list. Consequently, the application of memory models that assume latent rehearsals typically make use of overt rehearsal data. However, these data are not always available in some settings where the application of memory models has proven useful (e.g., clinical assessments of memory performance). In this paper, we show Bayesian statistical methodology can be used to infer the latent pattern of rehearsals needed to successfully apply a temporal model of memory to a clinical data set. We discuss the relevance of this research for those interested in neuropsychological assessment as well as cognitive psychologists interested in basic memory research.

Keywords: Alzheimer's disease and related disorders; Cognitive psychometrics; Hierarchical Bayesian modeling; Human memory; Missing data

Introduction

Quantitative models of human memory often rely on assumed latent memory processes. These assumptions are common to a range of memory models, based on different theoretical motivations (e.g., two-store vs. temporally based accounts of memory), and are used to account for a similarly diverse range of observed memory phenomena (for an overview, see Norman, Detre, & Polyn, 2008).

In general, however, these memory models are typically developed to account for data collected in the environment of a controlled laboratory experiment, and problems can arise when the model is forced to leave this environment and account for data collected in less controlled settings. These problems that should be considered by the developers and users of memory models, since they help determine the effectiveness of the model as an explanatory tool. In this paper, we outline one such problem that arises in the context of applying a popular memory model to clinical data coming from the diagnosis and assessment of Alzheimer's disease and related disorders (ADRD). The issue is that overhert rehearsal times that are often collected in the laboratory are not available in this clinical setting. Instead, we show how Bayesian statistical methods can be used to *infer* these rehearsal times from the available behavioral data.

The plan of the paper is as follows. In the next section, we provide an overview of our clinical memory data and the database from which they are obtained. Following this, we present the details of the memory model we use to explain these data and as well as the Bayesian statistical methodology used to connect our model to our data. It is shown that the basic version of this model, which does not include rehearsal processes, is unable to account for our data, agreeing with current results in the memory literature. This leads to a modification of the model that allows for the unkown rehearsal times needed for the successful application of the model to be inferred from the data. In addition, we show that the model can be fitted easily to more complex data sets than are typically used in previous applications. The results of fitting this model to this more complex data set, including inferences about the latent patterns of rehearsal, are then presented. We conclude with a discussion on the limitations of the current approach and suggest potential ways to improve our results, and we also discuss the relevance of this research for clinical applications and for cognitive psychologists interested in basic memory research.

Task and Data

Our memory data are a subset of a large clinical ADRD database (e.g., Pooley, Lee, & Shankle, in press). This database contains a wealth of information on thousands of ADRD patients—and often on their caregivers as well—who visit neurology clinics for dementia screening and assessment. Among other things, this information includes demographic information and information concerning personal medical history. In addition to this information, this database also contains the results of various psychological tasks that are administered as part of the cognitive portion of these dementia assessments. Of these numerous psychological tasks, however, we focus exclusively on a sequence of four free recall memory tasks, and we limit our focus to the data of 541 "cognitively normal patients" (i.e., those individuals judged not to have a form of ADRD by a trained clinican). In this sense, we are treating the data as standard memory data coming from normally functioning adults.

Collectively, these four memory tasks constitute a single multitrial free recall (MFR) task. Stimuli for this MFR task consisted of words based on the CERAD (Consortium to Establish a Registry for Alzheimers Dis-

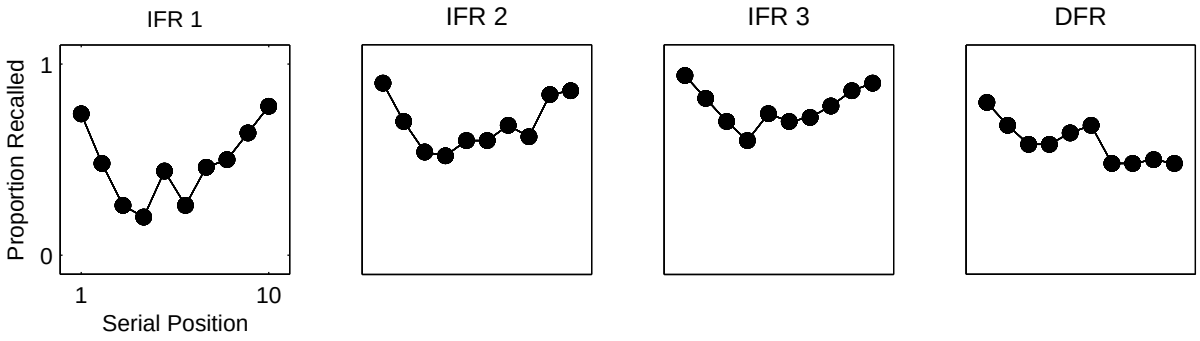

Figure 1: Serial position curves for multitrial free recall data. The "IFR i" panel shows the serial position curve for the ith immediate free recall task, and "DFR" panel shows the serial position curve for the delayed free recall task.

ease) word list (Morris, Mohs, Rogers, et. al, 1988), which serves as the basis for the neuropsychological portion of numerous ADRD assessments. These words, which included a mixture of common nouns (e.g., BUS, WEATHER, etc.), were chosen via statistical methods such as latent semantic analysis (Landauer & Dumais, 1997) with the goal of minimizing item effects such as semantic associability, differential item frequency, and so on.

Based on these stimuli, the MFR task was administered according to the following protocal. First, patients are presented with a study list of ten words. Following this presentation of the study list, the patients are asked to recall, in any order, the words on the list. Immediately following the completion of this first immediate free recall task, the same procedure is then repeated twice, with the same study list in the same order on the second and third immediate free recall tasks. After the third immediate free recall task, there is a delay during which the patients complete a variety of unrelated cognitive tasks as part of their dementia assessment. Following these tasks, there is a surprise delayed free recall task, in which the patients are asked to recall, in any order, the words on the previous study lists.

Since each patient produces a binary string as data, indicating whether or not a given word in a given serial (or input) position on the study list was recalled, it is often helpful to reduce the data and provide a group aggregate of recall performance. This aggregated data, averaged over patients for each serial position, is shown in Figure 1. These data demonstrate the standard serial position curve in free recall (e.g., Murdock, 1962), where words presented in early and late portions of the study list are better recalled than are words presented in the middle portion of the study list.

A Temporal Model of Memory

One goal of this research is to find a psychological model of memory that has the potential to be applied usefully to ADRD memory data in a clinical context. Serial position curves have been well studied in the memory literature; consequently, many theories and models of this curve have been developed.

Psychological models of memory typically take one of two forms. Two-store (or "buffer") models of memory (e.g., Raaijmakers & Shiffrin, 1981) treat the memory system as being goverened by processes that vary according to the time scale of the to-be-remembered information (i.e., they distinguish short- vs. long-term memory). In contrast, temporal models of memory (e.g., Brown, Neath, & Chater, 2007) assume that all aspects of memory, regardless of the time scale of the to-be-remembered information, are goverened by the same processes (i.e., they do not distinguish short- vs. long-term memory).

So we have two styles of memory model, each with complementary strengths and weaknesses, that could potentially be applied to our MFR data. Since temporal models of memory tend to be simpler in their implementation (which is well suited to exploratory research such as this) and would appear to be easier to scale up to larger data sets (which is well suited to potential future clinical applications), we explore one representative and currently popular temporal model in the current application.[1]

The Basic Model

The representative temporal model of memory[2] we apply was introduced by Brown, Neath, and Chater (2007). This model assumes that each word on the study list is representated in memory as a simple logarithmic compression of the time since its last rehearsal by the patient,

[1] We stress that this choice should *not* be taken as an endorsement of temporal models as superior to two-store models in accounting for memory. More specifically, we do not fundamentally believe in the general superiority of any of the currently popular (or unpopular) memory models.

[2] A note on our terminology: Models in cognitive psychology often are referred to by an acronym. In the current case, the model is known as SIMPLE (Scale-Invariant Memory, Perception, and Learning). In the current paper, we have chosen to identify the model by what we fell is its most important structural feature for our purposes; namely, the use of a single temporal dimension, rather than two "stores" and associated control processes, to account for memory performance.

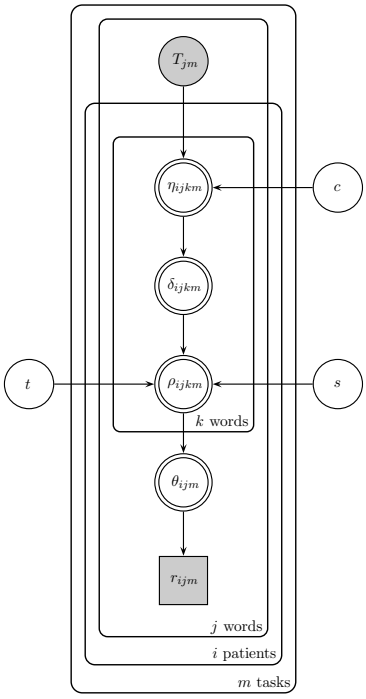

Figure 2: Graphical model representation of the temporal model for MFR data.

where $\log(T_j)$ is the representation of the jth word[3] in memory, which was last rehearsed at time T_j relative to the start of the recall task. Based on these representations, the *similarity* between the jth and kth words is given by

$$\eta_{jk} = \exp\left\{-c|\log(T_j) - \log(T_k)|\right\},$$

where the parameter c measures the "distinctiveness" of the memory representations. These pairwise similarities are then used to compute the pairwise *discriminability* between the jth and kth words, which is given by

$$\delta_{jk} = \frac{\eta_{jk}}{\sum_x \eta_{jx}}.$$

The *retrieval probability* for the jth and kth words is then calculated as a sigmoid function of the associated discriminability and is given by

$$\rho_{jk} = \frac{1}{1 + \exp\{-s(\delta_{jk} - t)\}},$$

where the parameter t measures the "retrieval threshold" for the words and s the "noise" in this retrieval threshold.

Based on these retireval probabilities, an arbitrary function (i.e., one that is unmotivated by psychological concerns) is used to compute the *response probability* for the jth word, which is given by

$$\theta_j = \min\left(1, \sum_k \rho_{jk}\right).$$

Finally, these response probabilities are are used to generate the binary recall data

$$r_j \sim \text{Bernoulli}\left(\theta_j\right),$$

where $r_j = 1$ indicates that the jth word was recalled and $r_j = 0$ indicates that the jth word was not recalled.

Applying the Model to MFR Data

The generative process just outlined, extended to account for the full structure of our MFR data, is shown as a graphical model in Figure 2. Graphical models (for an overview, see Jordan, 2004) provide diagrammatic representations of statistical models in which the nodes of a graph correspond to random variables, and the edges between these nodes correspond to the various independence assumptions of the statistical model the graph represents, with children independent of all other nodes given their parents. Our notational conventions are as follows: Square nodes represent discrete quantities and circular nodes continuous quantities. Shaded nodes represent observed quantities and unshaded nodes unobserved quantities. Stochastic quantities are represented by nodes with a single border and deterministic nodes are by nodes with double borders. Finally, independent replications of portions of the graph structure are enclosed within rectangles, which are referred to as "plates" in the literature on graphical models.

We apply two variants of the above model. Our first model assumes words are rehearsed exactly when they are presented at study. It also assumes that all individuals share the same values for the psychological parameters c, s, and t. Thus, in the graphical model in Figure 2, each of these deterministic quantities is enclosed in plates corresponding to each patient $i \in \{1, \ldots, 541\}$ and word $j, k \in \{1, \ldots, 10\}$, and the final study times of the words (relative to the start of the recall task) are not enclosed in the plate corresponding to the patients (i.e., each patient rehearses the words using the same temporal schedule). Furthermore, the psychological parameters c, s, and t are not enclosed within any of the plates, which means that they are both shared between the patients and remain fixed across the four recall tasks.

Our second model differs by assuming that there are covert rehearsals of the words after they have been presented. There are no data giving these rehearsal timings, so they must be inferred from the *available* data.

Statistical Inference

As should be clear from the above discussion, knowledge of the temporal schedule of the rehearsals (the T_{jm} variables in Figure 2) is critically important for the functioning of the model. In typical psychological experiments applying temporal models, experimenters often

[3]For simplicity of exposition, we use the phrase "the jth word" to mean "the word presented in the jth serial position of the study list."

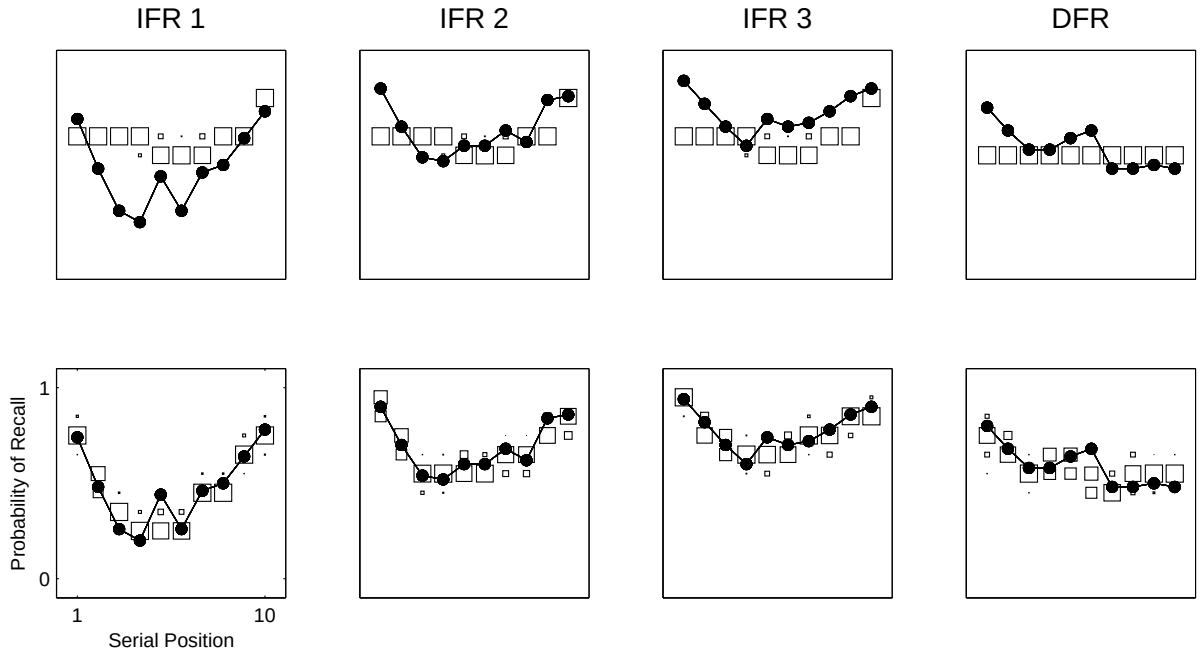

Figure 3: Posterior predictive distributions for the MFR task for two implementations of the temporal model, once in which rehearsal times are fixed (top panel) and the other in which they are inferred from the available data (bottom panel). The "IFR i" columns show the serial position curves for the ith immediate free recall task, and "DFR" columns show the serial position curves for the delayed free recall task. The black line shows the observed serial position curve for each task, and boxes represent posterior predictions made by the model, with the areas of the boxes proportional to the posterior predictive mass.

have some knowledge of these rehearsal times by having individuals rehearse the words out loud, and keeping track of the specific timing of the (observed) temporal characteristics of the rehearsals; thus, researchers using this experimental paradigm can apply this temporal model (e.g., Brown, Della Salla, Foster, & Vousden, 2007). Fortunately, this missing data problem can be addressed using Bayesian statistical methods, where our uncertainty about all unobserved quantities (including missing data such as latent rehearsal times) is expressed using probability distributions (for a comprehensive overview of these methods, see Gelman, Carlin, Stern, & Rubin, 2005).

The most basic conception of the Bayesian paradigm is quite simple: Start with a *prior distribution* for the unobserved quantities, condition on the observed data (in our case, the binary MFR data) to obtain the *posterior distribution* for these unobserved quantities (in our case, the psychological parameters and the latent rehearsal times), and use this posterior distribution to draw all the substantive conclusions of the analysis.

Choice of Prior Distributions The choice of prior distribution is quite important when using Bayesian methods. In the work presented here, we the same non-informative prior distributions for the latent psychological prameters c, t, and s as have been used in previous Bayesian applications this temporal memory model (e.g., Shiffrin et al., 2008). However, numerous logical constraints coming directly from the MFR task allow for a more informative prior on the latent rehearsal times, and in this paper we explore perhaprs the simplest of these logical contraints.

Consider the three immediate recall tasks. Patients are presented the words on the study list, one at a time (spaced approximately 2 seconds apart), and then asked to recall these words. Since individuals are allowed (in expeirments where the rehearsals are recited out loud) to rehearse the words at any time between their initial presentation and the start of the recall period, it seems reasonable to assume a uniform prior over the period of time from the presentation of any given word to the start of a given recall period, where this period decreases from the initial to the final serial positions. Since the fourth recall task is delayed and a surprise (i.e., it has no proper study period), it seems reasonable to assume that the prior distribution for the study time for each word is an identical uniform distribution extending from the start of the third immediate free recall task test period to the start the delayed free recall task. Of course, more realistic specifications of this basic idea can be made; however, it seems reasonable as a first pass approximation to a more complete prior distribution for the rehearsal times.

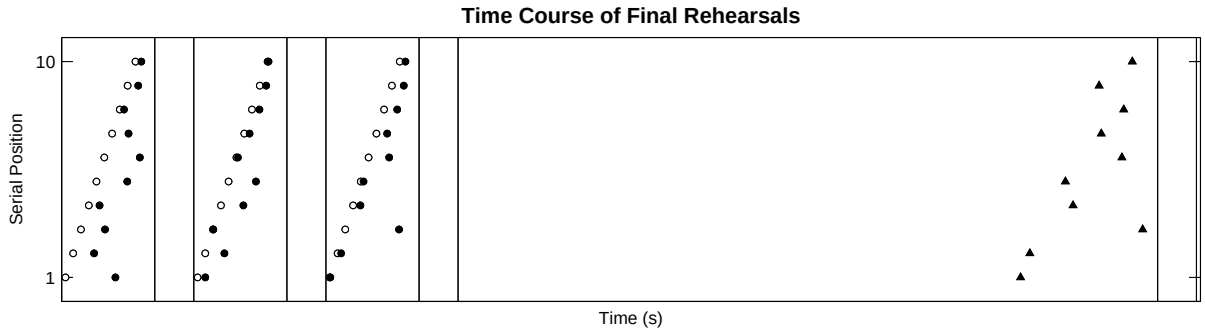

Figure 4: Temporal characteristics of the MFR task and the inferred rehearsal times for the MFR data. White markers correspond to stimulus presentations and black markers to MAP estimates of the rehearsal times; black circles represent rehearsal times for immediate free recall trials, and black triangles represent rehearsal times for delayed free recall trials. Solid vertical lines denote test periods of the MFR task when subjects are asked to recall words.

Modeling Results

Details on MCMC

In order to perform statistical inference, the graphical model shown in Figure 2 was implemented in WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000), software that uses a variety of Markov chain Monte Carlo (MCMC) methods (for an overview of such methods, see Gamerman & Lopes, 2006) to simulate the posterior distribution of the unknown quantities of interest in the model. Our results are based on samples from three MCMC chains, each consisting of 5,000 samples collected following a burn-in period of 1,000 samples. Convergence of the chains to the posterior distribution was assessed via the $\hat{R}$ statistic (Brooks & Gelman, 1998), which compares between- and within-chain variability.

Model Checking

Before we make posterior inferences about the quantitites of interest, we should check that our memory model is adequate. Many factors determine what makes a psychological model adequate, and just which of these factors are emphasized in a given analysis will ultimately depend on both the model itself and on the context in which the model is applied. In our current exploratory application, it is sensible sense to focus mainly on our model's descriptive adequacy (i.e., its to account for and describe interesting patterns in the observed data). In the Bayesian paradigm, this task is naturally accomplished using posterior predictive distributions (e.g., Gelman et al., 2004, pp. 165-172). Briefly, these distributions correspond to the data the model expects, based on the parameter values it has inferred from the observed data.

Figure 3 shows, for two implementations of the temporal model of memory discussed above, the posterior predictive distributions for four free recall tasks. The top panel shows the posterior predictive distributions for a model that fixes the rehearsal times T_{jm} to some fixed values[4], and the bottom panel shows the posterior predictive distributions for the full model, where these rehearsal times are inferred from the data. Each plot was generated by sampling parameters values (c, t, s, and the multiple T_{jm}) from the MCMC chains, and using these parameter values to generate serial position curves. The box sizes correspond to the amount of posterior predictive mass the model places on a given data point. Clearly, the model that allows the rehearsal times to be inferred from the data fits the data well while the model that assumes fixed rehearsal times *a priori* does not. Thus, we proceed to draw posterior inferences concerning the latent rehearsal times only for the full model.

Inferred Rehearsal Times

Figure 4 shows the inferred latent rehearsal times for the data. In this figure, white markers correspond to stimulus presentations and black markers to MAP estimates of the rehearsal times; black circles represent rehearsal times for immediate free recall trials, and black triangles represent rehearsal times for delayed free recall trials. Solid vertical lines denote test periods of the MFR task when subjects are asked to recall words.

Although there appear to be general patterns in these inferences, without additional constraints (from either theory or data) drawing substantive conclusions about the rehearsal times is difficult. Finding patterns in these rehearsal times, however, is not our goal here. Rather, our point is to demonstrate that the application of temporal models such as the one applied here need not be limited by the lack of relevant data, which, it is important to recognize, will typically be missing in ADRD settings. Thus, we feel that these results justify further exploration and extension to the full structure of the data described above. Once this is done, comparisons between the psychological parameter values learned for the

[4]Our investigations suggest that the exact pattern of latent rehearsals is underdetermined by the model and the data. However, the point here is that without the assumption of covert rehearsal, the temporal model cannot fit the data.

different stages can meaningully be made. As it stands now, however, we can only claim that these results show that our approach is in principle a sensible alternative to not applying a model to a given set of data.

Discussion

In this paper, we used Bayesian methods to apply a temporal model of memory to a subset of a clinical data set concerning the memory performance of ADRD patients. Critically, the Bayesian methods facilitated the application of the model when the key data needed to make the model work are missing from the data set. Furthermore, this is, to our knowlege, the first time that this specific temporal model of memory has been applied to the data from a full MFR task. Obviously, however, the model used here will need to be improved for fututre applications. For example, it was assumed that patients exhibit no individual differences in terms of memory performance. This is false when considering the case of normal aging adults due to ADRD, and even more clearly false when comparing the memory performance of these normal aging adults to groups of cognitively impaired individuals with some form of ADRD. Extending the model in such ways is straightforward using hierarchical Bayesian methods (e.g., Pooley, Lee, and Shankle, in press).

In addition to fixing these misspecifications at the level of the psychological model (used here for simplicity), we feel that the constrained nature of the MFR task presents an excellent opportunity to explore the use of prior distributions that are quite informative. For example, numerous physcial constraints determine when it is logically possible to rehearse an item presented in a given serial position on the study list. Our modeling here used perhaps the simplest formulation, and there exists additional information concerning the task to make this specification more realistic. Hopefully, such an improved specification would further improve the performance on the model.

Finally, a word about the potential users of this Bayesian methodology for memory research: It is easier for the experimental psychologists to perform experiments that than to learn the details of Bayesian methodology, so the why should an experimental psychologist care about this research? Our answer is that, even with the overt rehearsal paradigms currently used by experimental psychologists, the obtained data is only an approximation of the true rehearsal schedule going on in an individual's mind. Thus, overt rehearsal data can and should be used to constrain further the prior distributions on the rehearsal times. In contrast to basic research in experimental psychology, changes in experimental design are often hard (or costly) to implement on a large scale in clinical settings, and the existing data needs to be analyzed, in any case. Thus, it is in this area that we feel this work could yield the largest benefits. We feel that the modeling presented in this paper, although preliminary, are a positive first step in this direction.

Acknowledgements

This research was supported by award NIRG-08-90460 from the Alzheimer's Association.

References

Brooks, S., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics, 7*, 434-455.

Brown, G. D. A., Della Sala, S., Foster, J. K., & Vousden, J. I. (2007). Amnesia, rehearsal, and temporal distinctiveness models of recall. *Psychonomic Bulletin & Review, 14*, 256–260.

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

Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Jordan, M. I. (2004). Graphical models. *Statistical Science, 19*, 140-155.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*, 325-337.

Morris, J. C., Mohs, R. C., Rogers, H., et al. (1988). CERAD: Clinical and neuropsychological assessment of Alzheimer's disease. *Psychopharmacology Bulletin, 4*, 641-652.

Murdock, B. B. (1962). The serial position curve in free recall. *Psychological Review, 64*, 482-488.

Norman, K. A., Detre, G. J., & Polyn, S. M. (2008). Computational models of episodic memory. In R. Sun (Ed.), *The Cambridge handbook of computational psychology*. New York: Cambridge University Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93-134.

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