Posterior distribution analysis of the retention of briefly studied words

Lee Averell (Lee.Averell@newcastle.edu.au)
School of Psychology, University of Newcastle.

Andrew Heathcote (Andrew.Heathcote@newcastle .edu.au)
School of Psychology, University of Newcastle.

Abstract

The way in which memories decline over time has been studied for over a century, but remains an unresolved issue because of a lack of sufficiently precise data and effective model selection techniques. Here we address this gap with a particular focus on whether the decline in memory retention is complete or asymptotes above chance performance. We collected stem cued recall and stem completion data with tighter control over levels of interference and more data per retention interval than most previous studies. We analyzed the data using Hierarchal Bayesian models free from assumptions regarding the functional form of the forgetting curve. Population distribution estimate of retention at one hour well above the chance completion rate provided strong evidence for the use asymptote parameters in models of retention regardless of the functional form of the model.

Keywords: Forgetting; posterior likelihood; Long term memory; Recall.

In his analysis of the quantitative form of forgetting Wixted (2004a) dismissed the possibility of permanent retention (i.e., an asymptote) in forgetting functions. In contrast, Chechile (2006) supported asymptote parameters as effective in describing forgetting data over varying timeframes and paradigms (e.g., McBride & Dosher 1997; Rubin, Hinton & Wenzel, 1999), describing the omission of an asymptote parameter as a “serious failing”(p. 36). This paper investigates the whether parameters that represent permanent or very long lasting memories are feasible in both explicit and implicit memory. We begin by looking at the relevant literature on retention function modeling. We then describes an experiment measuring stem cued recall and stem completion, which was designed to have tighter control over both retroactive and proactive interference than previous studies, as well as obtaining more data per retention interval (lag) per participant over a greater range of lags. Our analysis of this data side-steps the issue of the correct functional form of the forgetting curve (e.g., power or exponential) by using hierarchal Bayesian estimation with no assumptions about functional form. We examine whether the posterior distribution of accuracy for the longest lag is shifted away from chance completion and, therefore, whether the inclusion of asymptote parameters in retention functions is warranted.

Modeling forgetting with asymptotes

The informal observation that humans can remember well learned and/or meaningful stimuli for a very long time has received robust experimental support. Bahrick’s (1984) seminal data showed that people accurately recognize Spanish language word definitions learned at high-school up to 50 years later. Similar performance has been found for street names from childhood suburbs (Squire, 1989), television programs (Schmidt, Peck, Paas & van Breukelen, 2000) and Shakespearean scripts (Noice & Noice, 2002). A characteristic of all these studies is a forgetting curve that becomes flat after some period of time and remains flat at above chance levels up to the longest tested retention interval. However, because all of these studies were cross sectional, and none controlled for rehearsal in the retention interval, they lack the rigor required for adequate inference about how individual memory traces are forgotten over time.

People can also retain information after only cursory study. However, the extended time course of forgetting has been less studied in this context; we review three influential studies that have examined retention over one minute to one hour. Wixted and Ebbesen (1991) studied free recall of words for lags between five and 40 seconds after study for either 2 seconds or 5 seconds each. The results showed that study time affected overall recall, with inferior performance for the shorter study time. Retention functions for both study times leveled off well above chance recall from roughly 15 seconds onward. Wixted and Ebbesen fit the data with a two parameter power function and a three parameter exponential that included an asymptote parameter and concluded in favor of the power function, not because the power function had lower least squares error, but rather because the exponential asymptote parameter produced estimates that they considered to be implausibly high.

Rubin, Hinton and Wenzel (1999) tested 300 participants on cued recall of paired associates after zero to 99 intervening trials, with the longest lag equal to around 10 minutes retention. They modeled the data with a sum of exponentials model that included an asymptote parameter (*a*3):

$$a_1 e^{-t/T1} + a_2 e^{-t/T2} + a3 \qquad (1)$$

T1 and T2 are rate parameters, where T2>T1. Based on reaction time data, which showed quicker responses for lags

0 and 1, Rubin et al. (1999) suggested that the first term $a_1e^{-t/T1}$ represented working memory. The remaining terms in the model represent either a single long-term process $a_2e^{-t/T2} + a3$, or an intermediate process $a2e^{-t/T2}$ and long term asymptotic level *a3* indicative of permanent or very long lasting memories. Regardless of the improved fit when using an asymptote parameter in their model, Rubin et al. were hesitant about the plausibility of permanent retention, saying "we believe the asymptote…represents a decline too small to detect in our experiments or even experiments with considerably longer delays" (p. 1173). They did, however, qualify this opinion by stating that the asymptote could plausibly represent a constant residual of context which serves to aid recall and would continue to produce above chance recall performance until the study context was totally different to the test context.

McBride and Dosher (1997) tested stem cued recall (an explicit memory task) and stem completion (an implicit memory task) between one minute and one hour after study. They showed that a power function including an asymptote parameter adequately captured the characteristics of both explicit and implicit data with only slight variation in the power function exponents. The asymptote parameter was employed because the data in both conditions was flat and above chance levels from about 15 minutes to one hour. When discussing the implications of the asymptote in their data McBride and Dosher (1997), like Rubin et al. (1999), were cautious, suggesting that "further decline would be measured in hours or days" (p. 380). In particular, they acknowledged a problem with the design, that study-test lag conditions were not evenly distributed throughout the experimental session. This may have produced variations in performance as a function of lag due to fatigue or proactive interference. Our experiment, which is modeled after McBride and Dosher's, controls this confound by equating the mean position within the experiment of each study-test lag condition.

Recently, Wixted (2004*a*) investigated the functional form of forgetting, with a particular focus on the use of above chance asymptotes in forgetting functions. Wixted modeled short term retention (Wixted & Ebbsen 1991), intermediate term retention (Rubin et al. 1999) and very long term retention (Bahrick, 1984) using data averaged across participants. Critically, Wixted examined the use of asymptotes by comparing a variant of the power model called the Pareto 2 model (Begg & Wickelgren, 1974, hereafter referred to as the Pareto model) with no asymptote and a three parameter exponential model that included an asymptote. Wixted found slightly better least squares fits for the Pareto than other models, leading him to conclude that asymptote parameters need not be included in forgetting functions.

However, Rouder and Lu (2005) showed that the loss of individual variability, as is inevitable when averaging across participants, leads to an asymptotic bias when modeling non-linear processes. In particular averaging across monotonic non-linear curves produces results that are more graded than curves used in the averaging process. The structural difference between the averaged curve and the individual curves that form the average need not be large for erroneous conclusions regarding the generating process to occur; Brown and Heathcote (2003) showed that exponential exponents need only differ by a single order of magnitude between participants for a power model to outfit an exponential model for averaged exponential data. Wixted (2004 a) acknowledged that such an averaging artifact may have influenced his analysis and conclusion.

The above review shows that the opinions as to the best quantitative description of forgetting are mixed. While many data sets show a leveling off at above chance accuracy, most researchers are reluctant to accept this trait as indicative of the underlying cognitive process being measured, citing methodological aspects such as a short measurement period as the cause. Further, as noted above, questionable analysis techniques (e.g., fitting to data averaged over participants) may have also confused this issue. What is needed is a technique that takes account of variation among participants, hierarchal Bayesian modeling is one such technique

Bayesian Analysis

Recent advances in Markov Chain Monte Carlo (MCMC) techniques have allowed Bayesian analysis to be applied to problems that were previously computationally or analytically intractable. A Bayesian analysis starts with "prior" distributions for parameters in a model. Prior distributions can be thought of as defining knowledge about parameter values before the data are observed. A parameter for the probability of retention at a particular lag, for example, might be given a uniform prior over the 0-1 interval to indicate that the parameter must be bounded between 0 and 1 (as it is a probability), but that within this interval all values are equally likely a priori. Importantly for our application, if we instead estimate the probit (i.e., inverse cumulative normal or "z" transform) of a probability parameter, a standard normal prior on the probit scale corresponds to a uniform prior on the probability scale.

A major advantage of Bayesian methods is that they make it practical to fit "multi-level" or "hierarchal" models (Rouder & Lu, 2005). A hierarchical analysis avoids the problems associated with averaging data, but is still able to model group data patterns, by assuming participants are drawn from a distribution. Each participant corresponds to a set of probability parameters (one for each lag) produced by a random draw from the distribution. The participant-level distribution is characterized by "hyper-parameters" corresponding to population estimates of the probability of retention at each lag. Hyper-parameters can also be estimated to account for correlations amongst retention probabilities at each lag. Correlation hyper-parameters can model for data where, for example, higher retention at one interval is associated with higher retention at other intervals (e.g., due to differences in participant's overall mnemonic ability).

We allowed for possible correlations by assuming each participant's set of seven probit transformed probability parameters (one for each lag) was drawn from a multivariate normal distribution with an arbitrary variance-covariance matrix. In a Bayesian hierarchical analysis priors need only be specified at the level of hyper-parameters. We assumed standard normal priors for the seven hyper-parameter means (and hence a uniform prior on the probability scale) and a Wishart prior over the variance-covariance matrix (main diagonal =1, off diagonal=0; Rouder, Lu, Sun, Speckman, Morey & Naveh Benjamin, 2007). Each participant's data (i.e., counts of correct recalls) was modeled by random draws from binomial distributions with probability parameters given by the participant-level distribution.

In summary, our Bayesian analysis allowed us to explore the use of asymptote parameters in forgetting functions without requiring a commitment to the particular functional form of the forgetting curve. In particular, we assessed the need for an asymptote parameter by investigating whether retention changes, and whether it remains above chance, over the longest lags. Because this hierarchical analysis takes account of variation at both the data and the participant levels, as well as correlation at the participant level, it produces interval estimates (called "credible intervals" in Bayesian analysis) for group parameters that properly reflect all sources of error. Hence, such intervals are realistically wide, and so provide a rigorous test of asymptotic performance.

Method

Participants.

Thirty two University of Newcastle students took part in the experiment. All were self-reported competent English language speakers. Participants received $35 to reimburse expenses incurred due to taking part in the experiment. The 32 participants were allocated into either an explicit (n=16) or implicit (n=16) condition.

Design

A pilot study was conducted to determine the chance completion rate of test words. Twenty participants were asked to complete 905 three letter word stems with the first four, five or six letter word that came to mind without any previous exposure to study material. All 905 stems had four or more possible completions with a maximum of 6 letters. Study words corresponding to each stem (critical words) were selected on the basis of natural language word frequency (based on the CELEX English corpus, Baayen, Pipenbrock, & van Rijnand, 1995) to have the second highest frequency of possible completions for the stem. Where two words were equal in frequency one was chosen at random. Of the 905 critical stem/word combinations the 786 words with the lowest completion probability (i.e. the words that were completed least frequently with the pre-selected target completion in the pilot study) were chosen to be the critical test set for the experiment. The pilot study gave a chance completion probability for the 786 words of 5.6%.

The main experiment lasted for 2.08 hours and was divided into two sections. Section one lasted for 62.4 minutes. It included 16 study-test cycles. Study cycles consisted of 17 pairs of study words which appeared in white on a black background at either side of the center of the screen. Test cycles consisted of 26 three letter word stems per cycle, which appeared one at a time in white on a black background with three trailing underscores following the last letter. Following a break of 7 minutes 48 seconds (equivalent to exactly two study-test cycles), section two commenced, which involved 14 study-test cycles and lasted 54.6 minutes.

The experimental materials consisted of 1020 study words (30 sets of 34) and 780 test stems (30 sets of 26). The 1020 study words consisted of the 786 critical words and 224 words drawn from an additional set of 642. The 780 test stems consisted of 546 of the 786 critical set stems as well as the remaining 119 stems from the pilot study and 115 filler stems that did not match any studied word. Each set of study words and test stems were drawn randomly from a word bank that included both critical and non-critical words, with the constraint that each set had the correct number of critical and non-critical words.

Retention was measured at seven approximately exponentially spaced lags (1.27, 2.63, 5.85, 9.75, 17.55, 33.15 and 64.35 minutes). The first two lags (1.27 and 2.63 minutes) were within-cycle lags and tested retention of words in the just presented study list. Test items for these two lags occurred, on average, 25% and 75% of the way through the test list. The number of within-cycle items tested in each cycle was either 1,2,3,4 or 7. When four within-cycle items were tested in the same cycle, tests were performed sequentially over test positions 5-8 and 18-21 respectively and the middle interval of these positions (6-7 and 19-20) when two items were tested. Where the number was odd the corresponding ranges were randomly selected from the pairs 5-7 or 6-8 and 18-20 or 19-21. In the test cycles where either lags one and two were tested 23 of the test cycles tested an even number of critical words (13 for lag one and 10 for lag 2)and 33 (15 for lag one and 18 for lag two) tested an odd number of critical words.

The other five lags were tested between study-test cycles, measuring retention intervals from 1 to 16 cycles in length. These tests occurred symmetrically and in the minimum interval around the middle of the test list (position 13.5), excluding within-cycle test positions. When testing multiple between cycle lags in a list, the test items from different lags were distributed randomly inside the interval. Critical words were allocated to lag conditions randomly but so that the average word completion probability, as dictated by the results of the pilot study, was as close to equal as possible across lag conditions. The average midpoint of study-test intervals was equated across lags to within .17 of a second of each other in order to control the fatigue and interference confounds on lag effects which potentially confounded the measurement of retention curves in McBride and Dosher's (1997) experiment.

Procedure

The procedure was identical for participants in both groups except for the stem completion instructions. The study-test cycles began with 17 pair-rating trials (34 words in total), in which the participant was required to rate which, of a pair of words, occurred more frequently in their linguistic experience. Each pair appeared on the screen for four seconds before the next pair appeared. The pair ratings task was used to insure that participants employed a consistent encoding strategy. Following the study list participants performed a stem completion task. Each three letter stem and three trailing underscores stayed on the screen for six seconds, during which time the participant was required to type a response. Participants in the explicit condition were instructed to try to complete the stem with a four, five or six letter word corresponding to a word previously seen in the pair-rating task. They were told that certainty was not necessary, and that if they were not sure they should guess. Participants in the implicit condition were told to complete the stem with the first four, five or six letter word that came to mind. All participants were forewarned not to pluralise a stem that was also a word by adding an 'S' at the end (e.g., CAR_), but that they could use 'S' to create a new word (e.g., BAS_). In the implicit condition participants were told that they should not respond with the plural of the stem if it was the first word that came to mind, but rather they should think of another word. The participants were also instructed to avoid slang or jargon, but that the use of proper nouns was permissible. They could use corrective keys such as backspace and delete when entering a response, so long as it was within the six seconds.

Results:

WinBUGS (Lunn, Thomas, Best & Spegelhalter, 2000) was used to obtain a single chain of 100,000 independent iterations from the posterior after discarding the first 25,000 iterations and only accepting every 150th iteration. Visual inspection of the chain confirmed convergence, and independence was confirmed by inspecting autocorrelation plots. The prior distribution for the probit transformed probability of completion was assumed to be a standard normal at each of the seven lags; however we found that posterior estimates were consistent across a normal prior distributions with larger standard deviations (i.e, 2 and 5).

Figure 1 shows the population posterior mean estimates (indicated by the circles) and the 95% credible intervals (error bars) for study completion probability at the seven lags. The credible intervals represent the range between the 2.5% and the 97.5% quantiles of the posterior samples. The solid horizontal line indicates the 5.6% chance completion rate. Participants in the explicit condition generally performed better than those in the implicit condition. Consistent with studies discussed previously, performance in both conditions declined monotonically for the first 15 minutes before leveling off well above chance completion.

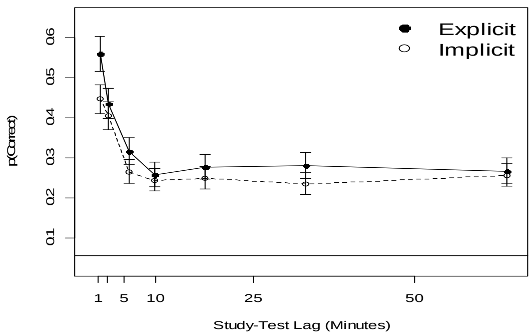

Figure1, Probability of correct completion as a function of lag for both explicit (solid line) and implicit (dashed line) conditions. Error bars represent 95% credible intervals. The solid horizontal line at the bottom of the figure represents chance completion rate.

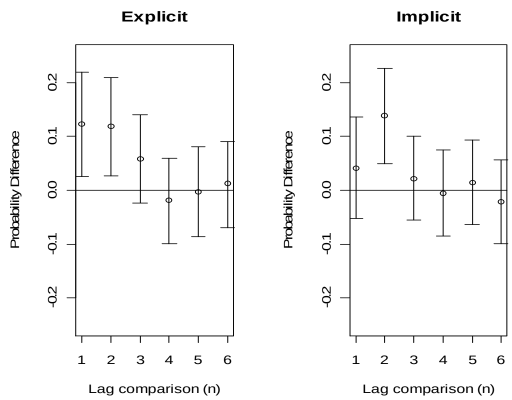

Figure 2, Posterior sample difference distribution for lag (n+1)-lag (n). Bars represent 95% credible interval of the difference distribution.

Figure two shows the mean differences between posterior samples from adjacent lags. The error bars represent the 95% credible interval for the difference distribution. When the error bars in figure two do not cross zero it indicates that there is a reliable difference in the posterior estimates of completion probability between adjacent lags. Note that these credible intervals take account of the correlations between adjacent lags, which were all positive with the exception of lags six and seven in the implicit condition, which were slightly negative. The plot shows that for the explicit condition there was a reliable decrease in completion probability from the first to the second lag, and again from the second to the third lag; thereafter no difference was reliable. In the implicit condition only the second and third lags were reliably differently. Of particular interest are the last two differences in each condition. They show that retention was stable among the fifth to seventh lags. The slight drop in the explicit condition between the sixth and seventh lag (mean difference estimate slightly above zero) and a slight rise in retention in the implicit condition (mean difference estimate slightly below zero) though both well within chance.

Figure three shows the posterior distribution for population retention probability at the longest lag (lag seven) for both the explicit and implicit conditions. The dashed lines represent the 95% credible interval for both distributions (explicit 95%CI=.2-.34, implicit 95% CI=.2-.31). In both conditions the chance completion rate was well below the 2.5^{th} percentile. In fact, the .001 percentile for completion probability in both distributions was 0.142, more than two times larger than the probability of chance completion.

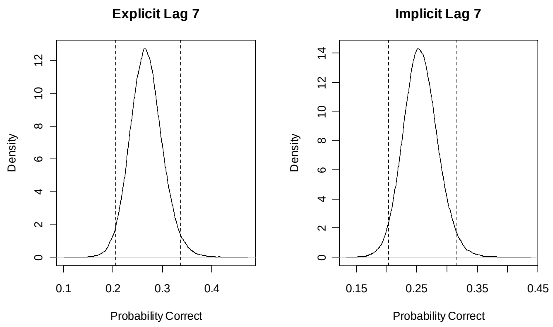

Figure 3, Posterior distributions for mean population study completion probability at lag 7 for explicit (left) and implicit (right)

Discussion

The hierarchal Bayesian analysis of this data shows that the population distribution for probability correct in the last lag of both the explicit and implicit conditions was well above chance. Moreover, performance was stable at this level from about 15 minutes on in both conditions. Given the cursory nature of the study these results strongly suggest that the use of an above chance asymptote parameter in any function used to describe this data is warranted.

The result is counter to the theory proposed by Wixted (2004ab) that memories ultimately decay completely. Wixted suggests that the eventual complete degradation of memory traces is due to the build up of retroactive interference that has ruinous effects on memory consolidation processes. Although the current study does not explicitly test this hypothesis, the unchanging performance between 15 minutes and one hour implies that if the build up of retroactive interference has an affect on memory performance it does so only in the first 15 minutes after study.

A possible explanation of the result is that performance in this task is strongly supported by cues provided in the test environment. The provision of the first three letters of the critical word serve as a strong retrieval cue and, as pointed out in a meta-analysis by Smith and Vella (2001), retrieval cues, such as an item cue, especially aid recall in long term memory tasks. Further, Zeelenberg, Pecher, Shiffrin and Raaijmakers (2003) showed a boost in priming when retrieval cues are given to participants, which suggests that implicit performance in the current experiment may have been supported by retrieval cues.

That retrieval cues help to maintain long-term memory performance above chance could also account for the asymptote in McBride and Dosher (1997) data set, which offered strong item cue support. A retrieval cue hypothesis is also in agreement with Rubin Hinton and Wenzel's (1999) alternate account of the asymptote parameter in their data; that asymptotic performance in their experiment represents a residual of study context at test. In Rubin et al. the retrieval cue was the test items' paired associate. This can be considered to be a weaker retrieval cue than a word stem and as such does not provide as much support, leading to a lower probability of recall in the long term than performance in stem cued recall designs. Such an effect is seen when comparing of asymptote parameters estimates in the Rubin et al. data set (10%) to that in the McBride and Dosher data set (28% explicit, 24% implicit). It is, therefore, possible that performance in memory experiments is heavily reliant on the retrieval cues that constitute contextual overlap, such as environment and study item information, between study and test, and that retention will remain above chance while there is a residual of context remaining in the test phase. The retrieval cue account of the results of the experiment is juxtaposed to the account offered by Wixted (2004ab), suggesting that failure to retrieve, and not a breakdown in the consolidation process, is the main cause of forgetting.

In both the current data set, and McBride and Dosher's (1997) data set on which this experiment was based, there is a strong similarity between both explicit and implicit performance. This is suggestive a single system underlying performance in both conditions where differences in performance are dictated by task demands rather than different neurological substrates (c.f. Kinder & Shanks, 2001). It could, however, be suggested that the implicit condition did not provide a "processes pure" measure of the implicit memory system if participants were using explicit memory to complete the stems.

However, in a near replication of the experiment reported here we ran three conditions; an explicit condition, and implicit condition and a "speeded implicit condition", in which participants were asked to respond with the first word that comes to mind as quickly as possible. Participants received a "too slow" warning if the first key stroke was longer than 1.5 seconds after the presentation of the test stem. It has been previously argued that responses emphasizing speed limit the use of conscious processes (Wilson & Horton, 2002). The experiment tested retention between one minute and one month over 4 experimental sessions. The results for the explicit and implicit conditions in the first session were very similar to those reported above (see Averell & Heathcote, 2009). Importantly, the speeded implicit condition showed a very similar pattern of results to the implicit condition in the current experiment, Averell and Heathcote's implicit condition, and McBride and Dosher's (1997) implicit condition.

One possible weakness in the present design is that test items for longer lags were drawn from fewer study lists than test items for shorter lags. This may have increased performance for longer lags because test items from the same list provide a context that could facilitate retrieval (Howard & Kahana, 2002). However, Averell and Heathcote's (2009) experiment minimized this confound and found that performance remained constant, at above chance levels at the longest lag.

To summarize, the use of asymptote parameters in modeling retention has been questioned as a legitimate extension of forgetting functions. The analysis presented here shows that asymptote parameters are warranted as valid extensions of models of retention. The omission of asymptote parameters may add to the difficulty in settling the issue of the most adequate quantitative form of the forgetting curve. For instance, Rubin and Wenzel (1996) showed that the exponential function without an asymptote did not fit much better than a linear model when fit to 210 existing data sets. However, when Rubin et al. (1999) designed data collection processes to maximized the ability to distinguish among forgetting functions an exponential that included an asymptote parameter out-performed all comparison functions.

References

Averell, L. & Heathcote, A (2009). Long term implicit and explicit memory for briefly studied words. In A. Taatgen, & H van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society.* Austin TX: Cognitive Science Society. ISBN 978-0-9768318-5-3

Baayen, R.H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX Lexical Database, Release 2* [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

Begg, I & Wickelgrn, W. A. (1974). Retention functions for syntactic and lexical versus semantic information in sentence recognition memory. *Memory and Cognition, 2,* 353-359

Brown and Heathcote (2003). Averaging Learning curves across and within participants.. *Behavior research methods, instruments and computers, 35,* 11-21.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113,* 1-26.

Chechile, R.A.(2006) Memory Hazard Functions: A vehicle for theory development and test. *Psychological Review, 113,*31-56.

Howard, M,W & Kahana, M.J (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269-299.

Kinder, A., & Shanks, D.R. (2001). Amnesia and the declarative/non-declarative distinction: A recurrent network model of classification recognition and repetition priming. *Journal of Cognitive Neuroscience, 13,* 648-669.

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

McBride, D. M., & Dosher, B. A. (1997). A comparison of forgetting in an implicit and explicit memory task. *Journal of Experimental Psychology: General, 126,* 371-392.

Noice, T., & Noice, H. (2002). Very long term recognition and recall of well learned material. *Applied cognitive psychology,16,* 259-272.

Ratcliff, R. & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9,* 347-356.

Rouder, J.N. & Lu, J. (2005). An introduction to Bayesian Hierarchal Models with an application in the theory of signal detection. *Psychonomic Bulletin and Review,12,* 573-604.

Rouder, J.N., Lu, J., Sun,D., Specman,P.L., Morey, R.D., & Naveh-Benjamin (2007). Signal detection models with random participants and item effects. *Psychmetrika, 72,* 621-642.

Rubin, D.C., Hinton, S., & Wenzel, A.E. (1999). The precise time course of forgetting *Journal of Experimental Psychology, 25,* 1161-1176.

Rubin, D. C.,Wenzel, A. E. (1996) One hundred years of forgetting: A quantitative description of forgetting. *Psychological Review, 103,* 734-760.

Schmidt, H. G., Peck, V. H., Paas, F., & van Breukelen, G. J. P. (2000). Remembering the street names of ones childhood neighborhood: A study of very long term retention. *Memory, 8,* 37-49.

Smith, S. M., & Vella, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8,* 203-220.

Squire, L. R. (1989). On the course of forgetting in very long term retention. *Journal of Experimental Psychology: Learning Memory and Cognition, 15,* 241-245.

Wilson, D. E. & Horton, K. D. (2002). Comparing techniques for estimating automatic retrieval: Effects of retention interval. *Psychonomic Bulletin & Review, 9,* 566-574.

Wixted, J. T. & Ebbesen, E.B (1991). On the form of forgetting, *Psychological Science, 2,* 409-415.

Wixted, J. T. (2004 *a*). On common ground: Jost's (1897) law of forgetting and Ribots's (1881) law of retrograde amnesia. *Psychological Review, 111,* 864-879.

Wixted, J. T. (2004 *b*) The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55,* 235-269.

Zeelenberg, R., Pecher, D., Shiffrin, R.M., & Raaijmakers, J.G.W. (2003). Semantic context effects and priming in word association. *Psychonomic Bulletin and Review, 10,* 653-660.