A Continuous Source Reinstatement Model of True and Illusory Recollection

Brendan T. Johns,[1] Michael N. Jones,[2] and Douglas J. K. Mewhort[1]
bj23@queensu.ca, jonesmn@indiana.edu, mewhortd@queensu.ca
[1]Department of Psychology, Queen's University, Kingston, On.
[2]Department of Psychological and Brain Sciences, Indiana University

Abstract

In studies of false recognition, subjects not only endorse items that they have never seen, but they also make subjective judgments that they remember experiencing them. This is a difficult problem for most dual process models of recognition memory, as they propose that false memories should be based on familiarity, not recollection. We present a new computational model of recollection based on the recognition through semantic synchronization model of Johns, Jones, & Mewhort (2012), and fuzzy trace theory (Brainerd & Reyna, 2002). In addition to standard and false recognition, the model successfully explains multiple studies on both true and false recollection.

Keywords: Memory; semantics; false memory; recollection.

Introduction

In studies of false memory, subjects not only endorse items that never actually occurred as being studied, but they also make subjective endorsements that they consciously remember those items. This phenomenon is prominent in the Deese/Roediger-McDermott paradigm (Deese, 1959; Roediger & McDermott, 1995), where subjects are given study lists composed of words that are all semantically associated with a single critical lure (e.g. *nurse*, *hospital*, *medicine*, etc… for the critical lure *doctor*), and has been termed illusory (Gallo & Roediger, 2003) or phantom recollection (Brainered, Wright, Reyna, & Mojardin, 2001).

The existence of this phenomenon has been established across a number of different empirical tests. In the original Roediger & McDermott (1995) study, a remember/know procedure was used, where subjects were asked to respond *remember* if they had the conscious feeling that they had seen that item during study, and *know* if they could not remember it but had a feeling of familiarity. Overall, critical lures were endorsed as being studied at roughly the same rate as studied items. However, subjects made more remember than know judgments to the critical lures, suggesting that it is not just pure familiarity that is producing the false memories. This is a surprising finding because the majority of memory models predict that the opposite pattern should take place, where familiarity should be the dominant process in false memories. The implication is that it is not just the phenomenology of experiencing an item that is used in recollection, but word-level properties (i.e. semantics) also exert an influence on the process.

The role of phantom recollection has been explained by fuzzy trace theory (FTT; Brainerd & Reyna, 2002). FTT proposes that there are two types of traces laid down in memory: verbatim and gist traces. The familiarity component of FTT proposes that the majority of false memory effects occur due to overlap between the meanings contained in gist traces and the critical word. Phantom recollection within this framework proposes that activation of a gist trace can retrieve a verbatim trace. However, verbatim retrieval can be used for recollection rejection, where the retrieved verbatim trace mismatches with studied materials (Brainerd, et al., 2003).

Recently, a new computational model of recognition memory has been proposed, the *recognition through semantic synchronization* (RSS) model of Johns, Jones, and Mewhort (2012) that is theoretically motivated and mechanistically tied to the proposals of FTT. RSS is based on the premise that the basis of the representation of a model should be rooted in a representation learned from the linguistic environment. This is important primarily due to the finding that models of semantics construct different similarity distributions than are typically assumed by memory models (Johns & Jones, 2010). Hence, the word representations used by RSS, are used are constructed with a semantic space model. Unlike FTT, the original RSS proposes that only gist traces are stored, but that verbatim processing is accomplished by diverting focus away from semantic information.

However, the RSS model does not contain a component that can account for the contribution of phantom recollection. To accomplish this, a mechanism suggested by both empirical examinations into manipulating modality and source information in the DRM paradigm (e.g. Gallo & Roediger, 2003; Arndt & Reder, 2003) and that of recollection rejection, will be integrated to the model. For example, Gallo and Roediger had subjects study lists in either the visual or auditory modality. At test they found that subjects will endorse a critical lure when it is tested in the same modality as its studied items, but will reject it when it is not. The results of this experiment indicate that subjects are retrieving perceptual information based upon semantics.

These results suggest that one mechanism at play in phantom recollection is the association between the semantic features of a word with surrounding source information. Since in the DRM paradigm studied associates have overlapping semantic representations with critical lures, the critical lures will have many of their semantic features associated with the source of its studied items. At test, these source features are reinstated, causing the critical lures to be accepted if the test source is congruent.

The first section of this paper will briefly describe the original RSS model, and the second section will describe the learning and retrieval of source information mechanisms that are integrated into the RSS framework. Finally, we will apply the model to standard recognition memory experiments and false memory results.

RSS Model

A complete description of the RSS model is contained in Johns, et al. (2012), as are relevant simulations that support the model; hence, the model will be discussed briefly here.

Semantic Representation

In order to accurately explain false memory effects, it is necessary to have an accurate semantic representation. Although there are a number of different representation types available, we employ a simple representation based on pure co-occurrence. In this representation, semantic memory is assumed to be a sparse matrix, where the rows are words, and the columns are documents. A cell gets a value of 1.0 if a word occurs in a document, and value of 0.0 otherwise. These vectors were constructed from the TASA corpus.

List Encoding

An encoding of a study list is assumed to be similar to the proposals of fuzzy trace theory (Brainerd & Reyna, 2002), which proposes that gist traces are formed. A simplifying assumption is made in the RSS that all the items that occurred on a study list are added into a single composite, along with random context noise. In the RSS, this gist vector is constructed as:

$$M = \sum_{i=1}^{n} T_i + R \qquad (1)$$

where **M** is the composite memory vector for the list, $\mathbf{T}_i$ is a studied trace for word *i*, *n* is the number of traces studied, and **R** is the context noise vector. Studied traces are normalized to unit length prior to the sum. The context noise vector consists of uniform random noise between 0 and a parameter value. This is the resulting memory trace that is used to make an old/new decision.

Synchronization

Synchronization is the overall process by which RSS determines if a word occurred on the list. The process operates by sharpening positive information about a probe word in the memory trace, and simultaneously leaking negative information. *Sharpening* operates by iteratively adding probe information (the non-zero locations within the word's representation) back into the memory trace. The amount of information added is dependent on the similarity between the probe and the composite (assessed with a vector cosine). This value is then divided it by the current iteration. Hence, even though the cosine increases across iterations, the amount of increase is constrained by the current level of synchronization and iteration. This is constrained by the sharpening parameter (α), which determines the total amount of information that is added into the vector, and controls the importance of the sharpening process in synchronization. Sharpening is described with the following equation:

$$(P_i > 0)\ \ M_{k,i} = M_{k-1,i} + \left(P_i * \frac{\cos(M_{k-1}, P)}{k} * \alpha\right) \qquad (2)$$
$$i = 1, \ldots, d$$

Where **M** is the current memory vector, **P** is the probe vector, k is the current iteration, d is vector dimensionality, and α is the sharpening parameter.

Leakage operates by reducing the non-defining portions of the probe from the composite (i.e., elements that are zero in the probe vector). This is done by multiplying these locations by the leakage parameter (a value between 0 and 1), which reduces the magnitude of those locations across iterations. This is described with the following equation:

$$(P_i = 0)\ M_{k,i} = M_{k-1,i} * \delta \quad i = 1, \ldots, d \qquad (3)$$

where **M** and **P** are the memory and probe vectors, k is the current iteration, and δ is the leakage parameter.

Decision

The RSS uses two information sources in its decision process: similarity information to make a *Yes* response, and contradictory information to make a *No* response. This is based upon results that suggest that these two decisions are based upon different information (e.g. Johns, E. & Mewhort, 2002). These two sources are accumulated across time, and whichever exceeds a criterion first is the decision that is made.

The amount of contradictory information is assessed by determining the difference in the pattern of the probe and the memory vector. This is accomplished by computing the absolute difference between the defining portions of the probe and the corresponding locations within the memory vector, and dividing this summation by the magnitude of the probe. Formally, this contradictory count is computed as:

$$Cont = \sum_{i=1}^{d} (P_i > 0) \left| \frac{P_i}{\sum_{j=1}^{d} P_j} - \frac{M_i}{\sum_{j=1}^{d} M_j} \right| \qquad (4)$$

This returns a real value between 0 and 1 with 0 indicating that all of the probe information is contained in memory, and 1 indicating that none of the probe information is contained within memory. If the process is not efficient more contradictory information will be accumulated, which makes it more likely that the probe will be rejected.

The RSS-Source model

The underlying assumption of the current work is that both true and phantom recollection is at least partly based on the reinstatement of source information. In terms of phantom recollection, this is the reinstatement of the source of its studied associates. During the synchronization process, the values in the composite memory trace that become sufficiently activated will reinstate attached information. When this source information becomes sufficiently similar to the study context, a judgment based on perceptual information can be made (i.e. a remember judgement).

Source memory representation

We represent source with a Gaussian vector sampled from *N*(0,1). Source will occasionally be the combination of two vectors (e.g. source + a list context, in a list discrimination task) depending on the task being simulated.

The connections will not be completely interconnected, instead random connections between the co-occurrence representation and source information will be used. The use of random connections is a property of the sparse distributed theory of memory (Kanerva, 1988). The lack of total interconnectedness means that source information is not stored at every feature, but instead it is distributed across a word's representation (there will be 10 connections between each location in the gist vector to the source vector). Each of the connections is initialized to a random value. Source vectors had a dimensionality of 2,000.

The learning rule to update a weight between the semantic values and source values will be updated as:

$$\Delta w_{i,j} = S_j \quad (5)$$

where i is a location within a word's representation, j is a location within the source vector, and **S** is the source vector. When a word is studied, the random connections between the non-zero locations of the word's representation and the source vector are updated. The strength of the update is based on the current value of the source vector. These connections are carried over to test, and are used to reinstate the source.

In summary, a combination of two source vectors are associated with the semantic representation of a word: 1) the overall context vector, assumed to represent the perceptual characteristics of the surrounding testing environment, and 2) the source vector, assumed to represent the specific source that a word occurred in.

Source reinstatement

In order to make a source decision, the source of a word is continuously reinstated across the synchronization process. Reinstatement will occur based on the activation of the semantic content of the composite. When a location in the vector exceeds a certain parameter (the reinstatement parameter), the learned connection values between the composite and the source vector are added into the reinstated vector. As the synchronization proceeds, the reinstated vector will become more similar to the study source (if the word was studied, or the word's associates were studied). Based on the accumulated similarity, a source judgment can be made.

The reinstatement process is described with the following equation:

$$(M_i > r)\; S_k = S_{k-1} + w_{i,R(M_i)} \quad i = 1, \ldots, d \quad (6)$$

where **M** is the memory vector, **S** is the reinstated source vector, k is the current iteration, r is the reinstatement criterion, i goes through each item in the composite, $\mathbf{R(M_i)}$ is the set of random connections between the current composite location and the source vector, and d is the dimensionality of the composite. This equation simply adds in the connection weights between the composite and the source vector, for any location within the composite that accumulates enough information to be activated sufficiently (that is, exceeds the reinstatement parameter). The process will operate across iterations, with the reinstated source vector being carried over. If the vector reaches a similarity threshold then it can be accepted. If it does not than the probe will be rejected.

Parameters and simulation details

There are five parameters in the original RSS model (sharpening, leakage, context noise, and two decision criteria). Our new source component of the model adds three more: 1) source noise, 2) reinstatement parameter, and 3) the source criterion. Given the amount of data that the original RSS model is able to account for (15 simulations were reported), and the 5 simulations reported here, model is certainly not overparamaterized. In Johns, et al. (2012) the parameter sets for the RSS model were held constant. Given the different tasks and manipulations that are used to test recollection, this is not able to be done here. However, we attempted to keep as many parameters constant as possible in the following simulations. Each task was simulated 1,000 times in order to remove any variance from the results.

Discussion

This section describes a new model of source recollection built off of the foundations of the RSS model. The model gradually reinstates the source of a word based on learned associations between the semantic representation of a word and the surrounding source of that word during study. If the reinstated source is of high enough quality than a source judgment can be made. We will use source judgment and remember judgment interchangeably in this paper. A know, or familiarity-based judgment, will be made based of the similarity value from the original RSS model. An item will be rejected using contradictory information. While there is clearly more to the conscious recollection of a word than simply the reinstatement of its source, this is a beginning mechanistic account of the recollection process.

Simulations

The approach taken here will be to examine a number of different tasks that researchers have used to test recollection across both true and false memory paradigms.

Response-Signal Task

One of the key data types that has been used to examine the dynamics of source recollection is the response-signal task (Hintzman & Caulton, 1997), a standard task to examine speed-accuracy tradeoffs. In this task, a recognition memory probe is presented for a certain amount of time (e.g., 100-2000ms) and is then removed, and the subject is asked to determine if the probe was studied or not.

The result that will be simulated here is by Hintzman & Caulton (1997). In this study, subjects studied list item auditorily or visually, and repeated 1 or 3 times. For modality judgments, they found that repetition decreased minimal retrieval time and slowed the rate to asymptote. In addition, repetition also increased the discriminability of the non-tested items.

This experiment was modeled by setting the source criterion to a constant (at a value of 0.57), and the contradictory criterion to a linear function, where the value (set at 0.993) is multiplied by the current iteration. The sharpening and leakage parameters were set at a rather arbitrary 0.75, while the context noise parameter was set at 0.08, and the reinstatement parameter was set at 0.0095. List items consisted of 40 words (sampled from the Toronto word pool; Franklin, et al., 1982). Half of the items were associated with one source vector, and the other with a different source vector. All words were associated with an overall context vector. The test vector was the sum of the overall context vector and one of the source vectors. Only source judgments will be made, due to this being a source discrimination task.

The results of this simulation are displayed in Figure 1, which demonstrates that the model gives a good account of the data from Hintzman & Caulton (1997) with an R^2 =0.97, p<0.001. For congruent modality judgments, it was found that when a word was repeated 3 times it caused an increase in the minimum retrieval time (2 iterations). For words that were repeated once it required a greater number of iterations (3) to successfully retrieve a word's modality, similar to the data. This demonstrates that when a word is repeated, it allows the source vector to be reinstated at a greater rate, due to the ease of the synchronization. Words that were repeated also had a greater overall level of performance.

For words that were studied from an incongruent source, both repetition levels were initially rejected at the same rate. However, as the number of iterations increased, the model rejected words that were repeated three times at a greater rate. This shows that repetition allows for the source of an item to be reinstated, which allows for easier rejection of incorrect source words, similar to a recollection rejection process.

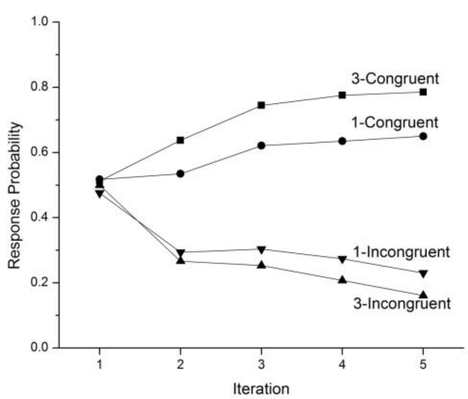

Figure 1. Simulation of Hintzman & Caulton (1997).

Remember/Know and Repetition

The first simulation demonstrated that the model can account for the effect of repetition in the response-signal paradigm; hence, it is natural to next test the model on the effects of repetition on remember/know judgments. Jacoby, Jones, & Dolan (1998) tested this empirically and found that across 1, 2, and 3 repetitions of a word in a study list, a corresponding increase in remember responses was found, but with a flat effect on know responses. Jacoby, et al. interpreted this result in terms of a dual-process model that has connections among its familiarity and recollection modules. The RSS-S model predicts this result to be due to an increase ease of reinstatement of the source of a word.

To simulate the effect, study lists of size 30 were used, with 10 being repeated once, 10 being repeated twice, and 10 being repeated three times. The sharpening and leakage parameters were set at, 0.55 and 0.255 respectively. The context noise parameter was set at 0.15, the reinstatement parameter was set at 0.035, and the source criterion was set at 0.69. The test vector was the context vector. Remember decisions were based on source similarity, while know responses were based on semantic similarity.

The results of this simulation are displayed in Figure 2. This figure demonstrates that the model is capable of accurately reproducing the results of Jacoby, et al. (1998). The rate of remember judgments systematically increases as a function of repetition, while the rate of know responses is unchanging. The quantitative fit of the model to that of Jacoby, et al. (1998; experiment 1) was R^2=0.99, p<0.001. The model can account for this result because repetition causes the associations between the word and the source to be strengthened, and also for the synchronization process to be increased in efficiency due to increases in familiarity. This results in the source being reinstated at a faster rate, and hence an increase in remember judgments.

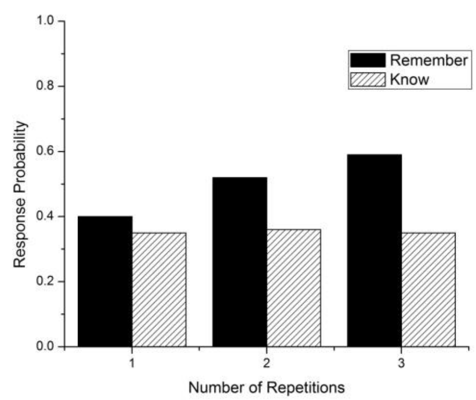

Figure 2. Simulation of Jacoby, et al. (1998).

Remember/Know Rates in the DRM Paradigm

In the classic DRM experiment of Roediger & McDermott (1995), remember/know rates to studied and critical words was measured, as well as old/new rates. They found that studied words had slightly higher rates of remember judgments, but critical words had a higher level of know judgments. Given that the original RSS model was able to give an accurate approximation to the levels of false memory to different list sets, it is key to establish that the RSS-S can give a good account to levels of remember/know.

To simulate this task, study lists of 4 DRM lists of 15 words each were selected from the lists of Roediger & McDermott (1995). Each word was associated with the same context information. As in Johns, et al. (2012) a new parameter set is used for false recognition simulations. The leakage and sharpening parameters were set at 0.5 and 0.85, respectively. The context noise parameter was set at 0.01, while the reinstatement parameter was set at 0.035. The similarity criterion was set at 0.994, the contradictory criterion at 4.0, and the source criterion at 0.56. The test vector was the overall context vector.

The results of this simulation are displayed in Figure 3. Studied items had higher rates of remember judgments, while critical words have higher rates of know responses. However, in both cases the levels of remember judgment's was greater than know responses, consistent with the data from Roediger & McDermott (1995). The fit of the model to the data was R^2=0.92, p<0.01. This simulation shows that the model is successfully associating the semantic structure of words with the source that it came from. This causes the model to have phantom recollection of critical words, since these word's representations were not directly associated during study. Instead, it was the cumulative association between the source vector and the word's semantic representation across the studied associates.

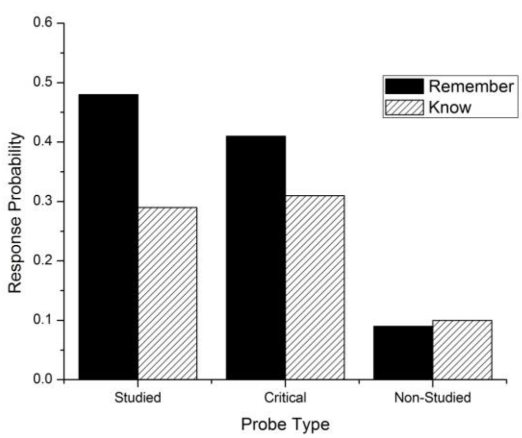

Figure 3. Simulation of Roediger & McDermott (1995).

Association of Source with Semantics

Arndt & Reder (2003) conducted a set of experiments where the correlation between source and semantics was manipulated, and tested with a remember/know task. The results of this experiment suggest that source is strongly connected to semantics during study, and hence is a key piece of empirical data for this model. In this study, subjects read lists of words that were associated with a critical item, similar to a typical DRM experiment. However, the font with which word appeared in was manipulated. There were three font conditions: 1) correlated: all associates to a critical word were studied in the same font, and this font was only used for that list, 2) uncorrelated: there were six fonts that were randomly assigned to study lists, and 3) unique: all studied items occurred in a different font. Studied items were tested in the font with which they were studied in. For the correlated condition, lure items were presented in the font that the list items were studied in, while in the uncorrelated condition they were presented in one of six fonts. New test words were assigned a random font.

For studied items, it was found that the uncorrelated condition slightly reduced the rate of remember responses, but this effect disappeared in the unique condition. The rate of know responses was constant across all three conditions. However, there were much larger effects for the critical lures. In the correlated condition, subjects made more remember than know judgments to critical lures, but not at the same level as studied items. The uncorrelated condition eliminated this effect, where the rate of remember and know responses were similar. This effect was attenuated in the unique condition, where the levels of remember responses decreased further, while know responses increased.

Table 1
RSS-S simulation of the results from Arndt & Reder (2003)

Font Condition	Item Type	Remember	Know
Correlated	Studied	0.63	0.25
	Critical	0.41	0.23
	New	0.04	0.08
Uncorrelated	Studied	0.61	0.24
	Critical	0.32	0.25
	New	0.06	0.06
Unique	Studied	0.58	0.26
	Critical	0.17	0.28
	New	0.02	0.09

In order to simulate this result, 6 DRM lists of 12 items were added into a study list, in the same manner as Arndt & Reder (2003). The lists were attained from the normed DRM list set of Stadler, Roediger, & McDermott (1999). Each font was assumed to be a unique source vector, and each word was associated with the summation of this vector with the overall context vector. In the correlated condition 6 unique vectors were used, 12 in the uncorrelated condition, and 72 were used in the unique condition. All parameters were kept constant from Simulation 4.3., with the exception of the source criterion being changed to 0.4. The test vector was the overall context vector summed with the source vector that was dictated by the empirical manipulation (described above).

The results of this simulation are contained in Table 1. This table shows that the model is able to give an excellent fit to the results of this experiment, with an R^2=0.95, p<0.001. For studied words, the level of remember or know responses was not overly impacted by the font condition. However, for critical words, there was a large effect. In the correlated condition, the model made more remember than know judgments to critical words, and this difference was highly reduced in the uncorrelated condition. Finally, this effect reversed in the unique condition with more know responses being made than remember responses. This simulation validates this approach taken: source becomes associated with the semantic representation of a word.

Effects of Association on Illusory Recollection

A common finding within the false memory literature is that as the number of associates that are studied is increased, a corresponding increase in the level of false recognition is also found. Gallo & Roediger (2003) tested how this affected illusory recollection by conducting a DRM experiment with source judgments. Critical words had 5, 10, or 15 associates studied, and their associates were studied in either the auditory or visual modality. At test, participants were presented with studied and critical words in a congruent or incongruent modality and were asked to determine if that was the correct source. For studied items, as the number of related items studied was increased a corresponding increase in the ability to accept the correct source was exhibited. There was no effect of number of associates on rejecting incongruent sources. For critical words, the probability of making a source judgment increased as the number of associates studied in that source increased. Additionally, the probability of rejecting an incongruent source also increased as the number of studied associates increased.

This result will be simulated by selecting 6 DRM lists from Stadler, et al. (1999). Two of the lists will have 5 studied associates included, two will have 10 associates studied, and two will have 15 associates studied. Each list will be associated with one of two source vectors (to represent the two modalities). Only the reinstatement parameter (set at 0.52) and context noise parameter (set at 0.004) were changed. All other parameters are consistent with Simulation 4.2. Given that only source judgments EW simulated here, similarity information was not used.

Table2
RSS-S simulation of the modality judgment results

Item Type	List Length	Congruent	Incongruent
Studied	5	0.59	0.19
	10	0.61	0.19
	15	0.63	0.18
Critical	5	0.52	0.29
	10	0.6	0.25
	15	0.64	0.19

Table 2 presents the results of this simulation, and demonstrates that the RSS-S gives a close correspondence to the results of Gallo & Roediger (2003) with an R^2=0.97, *p*<0.001. As the number of studied associates increased, the model produces a corresponding increase in the probability of accepting a congruent source increased for both studied and critical words. For critical words, there was also an increase in rejecting incongruent sources as a function of number of studied associates, suggesting that the greater level of connection between source and semantics allowed for more efficient recollection rejection. This demonstrates the level of association between the source of a word and its semantic representation a central aspect in both true and false recollection.

Discussion

This paper describes an attempt to integrate a recollection component into a mechanistic account of recognition memory. The RSS model of Johns, et al. (2012) bases its operation around a plausible representation of a word's semantic content, due to the finding that random representations (as used by the majority of memory models) construct different similarity distributions than semantic models (Johns & Jones, 2010). The model operates by forming a gist trace of a study list, and makes a recognition decision by attempting to synchronize a probe vector with this gist trace. The efficiency of the synchronization is based on the semantic overlap between the probe and gist, and the decision that is made is based on the success of this process.

References

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language*, 48, 1-15.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*, 164-169.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 307-327.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17-22.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). Norms for the Toronto word pool. Behavior Research Methods and Instrumentation, 14, 375–399.

Gallo, D. A., & Roediger, H. L. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition*, 31, 1036-1044.

Hintzman, D. L., & Caulton, D. A. (1997). Recognition memory and modality judgements: A comparison of retrieval dynamics. *Journal of Memory and Language*, 37, 1-23.

Jacoby, L. J., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgements and exclusion errors. *Psychonomic Bulletin & Review*, 5, 705-709.

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, 17, 662-672.

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2012). A synchronization account of false recognition. *Cognitive Psychology*, 65, 486-518.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 803-814.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create memories. *Memory & Cognition*, *29*, 424-432.