Are Random Representations Accurate Approximations of Lexical Semantics?

Brendan T. Johns (johns4@indiana.edu)
Michael N. Jones (jonesmn@indiana.edu)
Department of Psychological and Brain Sciences, Indiana University
1101 E. Tenth St., Bloomington, In 47405 USA

Abstract

A common assumption made by cognitive models is that lexical semantics can be approximated using randomly generated representations to stand in for word meaning. However, the use of random representations contains the hidden assumption that semantic similarity across randomly selected words is normally distributed. We evaluated this assumption by computing similarity distributions for randomly selected words from a number of well-know semantic measures and comparing them with the distributions from random representations commonly used in memory models.

Keywords: Memory models; semantics; episodic recognition

Introduction

A model of a cognitive phenomenon typically requires an account of both representation and process, and how the two interact (Estes, 1975). These two aspects of a model are interdependent, with the process requiring a representation on which to operate, and the representation requiring a process to simulate behavior. A common practice in cognitive modeling is to use randomly generated representations if the theorist wishes to evaluate a process mechanism, but is unsure of the correct psychological structure or features to use as a representation. This practice makes it unlikely that the representation is biased towards supporting the process model, and the process account can be later refined when further research reveals the correct representation. Over the history of computational modeling, emphasis has been placed on processing over representation.

If insufficient research exists to point towards the correct representation, random representations often provide a useful alternative or simulation of the process would be impossible. An excellent example is Hintzman's (1986) use of random representations to simulate schema abstraction using Posner and Keele's (1968) stimuli. Briefly, stimuli were random dot patterns, and exemplars of the same category were random perturbations of a prototype pattern. Without needing to account for how the human visual system represents dot patterns, Hintzman was able to create equivalent structure in his simulation by generating random prototypes and exemplars.

Random representations have been commonly used in models of episodic memory, for example, recognition, recall, and paired-associate learning. In global matching models of recognition memory (e.g., Hintzman, 1986; Murdock, 1982; Shiffrin & Steyvers, 1997) decisions are made by assessing the similarity of the probe word to the (usually noisy) study items with a particular processing and decision mechanism. The use of random representations in these models produces a hidden assumption that the distribution of similarity across randomly selected words is symmetric and approximately Gaussian.

The distributional assumption comes from the design of a typical memory experiment in which random words are used. In these experiments, random words are selected from a word pool (e.g., Friendly, et al., 1982). Because words are randomly selected, they are assumed to have only random similarity on dimensions extraneous to the experimental manipulation (e.g., orthography, phonology, semantics, etc.); however, this assumption is unlikely to be true. Hence, it is common to explicitly control extraneous factors such as frequency. In this examination, we focus on semantics—a factor often ignored because it is difficult to quantify and control. In assuming that two randomly selected words have only a random expected semantic similarity, random representations seem appropriate.

However, the use of these representations assumes that semantic similarity is randomly distributed across all sampled words. We demonstrate in the following analysis that this is unlikely to be the case with real words, and may produce consequences for conclusions drawn from process models that have used random representations.

Analysis

To evaluate the assumption of random similarity, comparison distributions are needed. Our analysis will utilize three types of semantic similarity measures to create distributions—similarity measures computed from: 1) free association data, 2) a hand-coded lexical ontology (WordNet), and 3) corpus-based co-occurrence models.

Semantic Measures

1. Word Association Space (WAS). Steyvers, Shiffrin, and Nelson (2004) developed a method for inferring semantic representations from free association data. Steyvers et al. represented the free association data for the 5000 cue words from Nelson, McEvoy, and Schreiber's (1999) norms in a word-by-word matrix, where each entry was the probability of a cue word (the row) eliciting the response (the column). This matrix was then reduced in dimensionality using singular value decomposition so that each word was represented by an abstracted 400-dimensional vector. Steyvers et al. demonstrated that the resulting vectors are a good predictor of similarity effects in recognition, recall, and other behaviors.

2. **WordNet Similarity**. WordNet (Miller, 1990) is a hand-coded lexical database encoded as a network in which nodes contain one or more synonymous words. These nodes are then linked together via different types of lexical relationships (e.g. hypernymy and holonymy) and based on these relationships it is possible to build a measure of semantic similarity between two given words using network statistics. A variety of methods that have proposed to do compute similarity, but the measure that seems to best map onto human similarity ratings is the Jiang-Conrath distance measure (JCN; Maki, McKinely, & Thompson, 2004). JCN is a network distance measure that basically counts the number of nodes and edges between two concepts in the database.

3. Latent Semantic Analysis (LSA). This method (and those that follow) differs from the WAS of Steyvers, et al. (2004) in that it does not use human behavioral data to create a semantic representation but, rather, uses statistical regularities computed from a large text corpus. In LSA (Landauer & Dumais, 1997), a word-by-document matrix is created by tabulating the frequency that each word occurs in a given document, inversely weighted by the word's marginal frequency and entropy over documents. The dimensionality of this matrix is then reduced using singular value decomposition so that each word is represented by a vector containing the 300-400 dimensions with the largest eigenvalues. Words that frequently co-occur in similar documents will be represented by similar vectors.

4. BEAGLE. In the BEAGLE model of Jones and Mewhort (2007), a distributed holographic representation of a word is built through experience with a text corpus. Words are initially represented by random Gaussian vectors, and a word's semantic representation is created by summing and convolving (cf. Murdock, 1982) other words that occur in sentences with a target word. The use of convolution allows order information to be included (the sentential position of the word relative to other words), as well as the basic co-occurrence information in LSA. This associative mechanism affords inclusion of rudimentary syntactic knowledge in the vector representation of the word.

5. The COALS model. Unlike the two previous models, COALS (Rohde, Gonnerman, & Plaut, submitted) is not designed to explain human learning, but rather to create a co-occurrence metric that yields the best predictions on a variety of semantic tasks. The model creates a word-by-word matrix, with modifications to how values within the matrix are computed (i.e. correlations are used instead of pure co-occurrence count). This large, sparse matrix is subsequently reduced in dimensionality with SVD in the same way LSA reduces a co-occurrence matrix.

6. Pointwise Mutual Information (PMI). PMI uses a pure co-occurrence count across a large text corpus to create a measure of similarity between two words (e.g., Recchia & Jones, 2009). As with COALS, PMI is not meant to be a model of human learning or representation, but rather a scalar measure of similarity between two words. PMI is essentially computed by taking the probability of observing word x and word y together and dividing by the probability of observing x and y independently. Recchia & Jones computed PMI values over a very large corpus of Wikipedia articles (approximately 400,000 articles), and found that PMI produced a significantly better fit to human rating data than LSA or other semantic similarity metrics.

Random Representations

To compare to the distributions created by the semantic measures, we explored five common types of random vectors that have been used to represent semantics in influential models of memory.

1. Random Gaussian Vectors. A word's representation is created by randomly sampling vector elements from a Gaussian distribution with a certain mean (typically zero) and variance (usually 1/N, where N is vector dimensionality). This type of representation has been used in a variety of models of recognition (e.g. Murdock, 1982), and recall, among others. In the following analysis, vectors were created as in Murdock (1982), with a vector size of 250, a mean of 0 and an SD of ($\sqrt{1/250}$).

2. Gamma Vectors. A word vector is created by sampling integers from a gamma distribution:

$$P[V = j] = (1-g)^{j-1} g, j = 1,...,\infty \quad (1)$$

Where g is a parameter between 0 and 1 that defines the environmental base rates for the different feature values. This type of representation has been used in the highly successful REM model of recognition memory (Shiffrin & Steyvers, 1997), and related models. We constructed these vectors as specified in Shiffrin & Steyvers (1997), with a length of 20, and a g = 0.45 (the parameter used to create high frequency words).

3. MINERVA vectors. In the influential MINERVA 2 model of memory (Hintzman, 1986), vector elements are assumed to be randomly selected from the set of {-1, 0, 1}. A value of 1 is intended to represent a positive link between the word and that feature, a -1 represents an inhibitory link, while a 0 is defined as either irrelevant or unknown for that particular word and feature. Vectors were constructed with a length of 20. Similarity for these vectors was calculated with the following equation:

$$s_i = \sum_{j=1}^{D} \frac{P_i \cdot T_{i,j}}{n} \quad (2)$$

Where D is the size of the vectors, P is the probe word, T is a studied memory trace and n is the number of non-zero items in P. The value is then transformed by cubing it.

4. Sparse Binary Vectors. In this type of distributed representation, the majority of entries are zero, with some entries having the value of 1 at random locations. For instance, in Plaut (1995) items in a word's semantic representation had a 10% probability of being non-zero. Sparse binary vectors have been used to model lexical priming (Plaut) and recognition memory (Dennis & Humphreys, 2001), among other domains. Similar to Plaut's

simulations we generated vectors with a length of 100 and each item having a 10% probability of being non-zero. In addition, binomial distributions (with a sparsity of 50%) will also be tested to examine the effect of sparseness on the similarity distributions.

5. Dichotomous Vectors. Another common type of representation used in connectionist modeling is a random vector composed equally of 1 or -1. These are similar to MINERVA vectors, but without any zero-valued elements. Dichotomous vectors have been used in variety of models, such as connectionist models of semantic priming (e.g., Masson, 1995). We use vectors with a length of 100 in the following simulations.

Method

To calculate similarity distributions using the semantic measures, 1000 words were selected from the Toronto word pool (Friendly, et al., 1982), and the similarity between each word in the pool was computed. Next, 50,000 of these semantic comparison values were randomly sampled to examine the distribution of similarity values. In the WAS, LSA, and BEAGLE models the similarity metric used was a vector cosine (a normalized dot-product), while in COALS Pearson's correlation was used.

For the randomly generated representations, we created a distribution of 100,000 similarity comparisons for each representation type. The distribution was constructed by randomly generating two vectors from the given representation type and computing the similarity between them. Similarity was vector cosine for all representations.

Toevaluate distribution shape, two different methods of assessing normality were employed: 1) skewness, and 2) normal quantile-quantile (Q-Q) plots. Skewness is the third moment about the mean, and signals asymmetry in a distribution. Q-Q plots are used to assess the difference between an observed distribution and a theoretical (in this case Gaussian) distribution. The standardized values of the comparison distribution are plotted against the respective values for the Gaussian, and any discrepancy signals a deviation from the theoretical Gaussian distribution.

Results

The skewness values for the similarity distributions of both the semantic spaces and random representations are displayed in Figure 1. As the figure shows, all the semantic spaces create positively skewed similarity distributions. That is, there tends to be a greater number of low similarity scores and a small number of high similarity scores in a given distribution of randomly selected words. Co-occurrence models (LSA, BEAGLE, and COALS) have the lowest skew (from 1.06 for BEAGLE to 2.01 for COALS). The PMI distribution produced the largest skew, likely due to the fact that this method does not abstract across documents, but is instead a pure co-occurrence count. Even with this shortcoming, PMI has been shown to be very effective in fitting human semantic similarity ratings

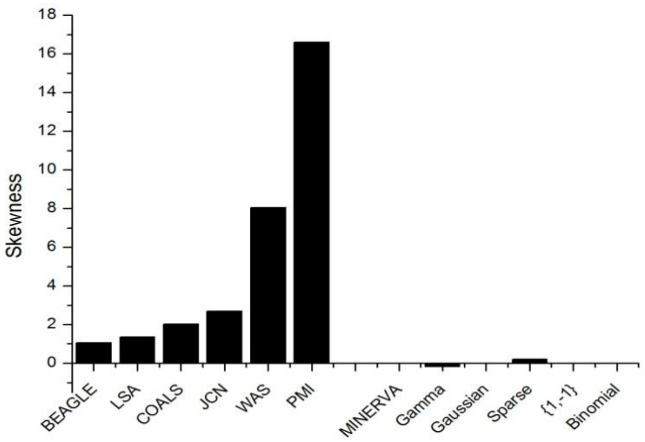

Figure 1. Levels of skewness for the different distributions.

(Recchia & Jones, 2009). In the middle was the JCN measure with a skewness of 2.61 and the WAS of Steyvers, et al. (2004) with a skewness of 8.04, which signals a highly skewed distribution.

In contrast, all of the random representations produced skewness values of essentially zero (this is expected by their construction). The only distribution that is mildly positively skewed is the sparse binomial distribution with a skewness of 0.21, while the Gamma distribution is actually mildly negatively skewed with a value of -0.17.

The Q-Q plots are displayed in Figure 2 for the semantic space distributions (left panel) and the distributions computed from the random representations (right panel). Due to space limitations, only 4 graphs were included, but these are diagnostic of the remaining distributions. Again, the semantic space distributions show significant deviation from the expected Gaussian distribution. Specifically, the semantic space distributions are skewed to the right, with all of the models having lower than expected number of large similarity values. They also tend to have greater than expected low similarity values. Again, the random representation distributions produce very different results—there is little deviation from normality.

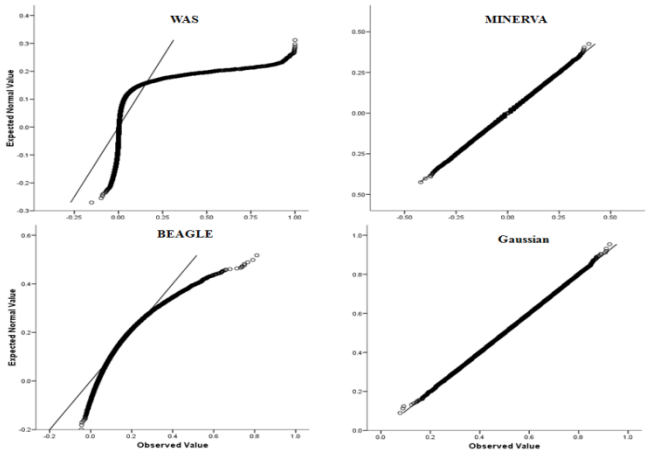

Figure 2. Q-Q plots for semantic and random vectors.

This simple analysis demonstrates that the similarity distributions created by semantic space models and randomly generated representations are considerably different. Two randomly selected words are likely to be less similar (relative to the other values in the distribution) for semantic models, than for random representations.

Demonstrations

In order to show the potential impact that the use of random representations may have, two simple demonstrations were conducted using data from recognition memory tasks.

Demonstration #1: Signal Detection Theory

The purpose of this demonstration is to show what effect skewed similarity distributions will have on a signal detection theory (SDT) based process, which is the dominant decision making process within recognition memory (Shiffrin & Steyvers, 1997; Dennis & Humphreys, 2001). In order to accomplish this, a recognition process with SDT is simulated by sampling from both skewed (semantic) similarity distributions as well as normal (random) similarity distributions. Recognition is then simulated by fitting an optimal criterion to separate old and new items, and the resulting d-prime values for the different distributions will be compared to behavioral results.

In order to compare the different similarity distributions, a normalization procedure was necessary. This was done by taking the distributions from each of the semantic metrics and random representations and normalizing them to have a range of 0 and 0.5 and a mean of 0.25. This procedure allows us to evaluate the shape of the distribution while centering the distributions on the same mean.

Evidence distributions for new and old items were simulated for lists of 20 words. The evidence for a probe was the similarity of the probe to the 20 items on the list. For 'new' probes, this evidence was simply the mean of 20 randomly sampled similarity values (as new probes are randomly similar to the contents of memory). For 'old' probes, this evidence was the average of the similarity of the item to itself and the other items on the list (simulated as the mean of 19 randomly sampled similarities and the value of 1, representing the similarity of the word to itself). This process was repeated 50,000 times for each similarity distribution.

To compare the resulting evidence values, the discriminability (measured with d-prime) was calculated for each simulation—d-prime is a measure of how distinct studied items are from non-studied items. Figure 3 displays the d-prime values for the different similarity distributions compared with the d-prime from a simple recognition experiment which used a list length of 20 (Dennis, Lee, & Kinnel, 2008). As the figure illustrates, all of the semantic distributions have higher d-prime than do the random distributions. In addition, the d-prime values for the random representations are much closer to the behavioral data from Dennis, et al. The difference in magnitude demonstrated for d-prime values for semantic and random similarity was statistically reliable, $t(11) = 4.75$, $p < 0.001$. To evaluate the effect of skew in the similarity distributions on the resulting d-prime values, we computed the partial correlation between d-prime and skewness (controlling for kurtosis and variance) for the distributions, which resulted in a robust $r = 0.913$, $p < 0.001$.

The skewness of the similarity distribution has a large effect on the calculation of evidence distributions because the probability of sampling lower similarity values is much greater than in a symmetric distribution. Hence, with 'true' semantic representations an old item tends to be more distinct from other random items on the list, producing a greater difference between old and new evidence distributions. This demonstration is certainly not meant as a refutation of signal detection theory, but instead demonstrates that using realistic representations of semantics will impose significant constraint on a processing model's ability to simulate data.

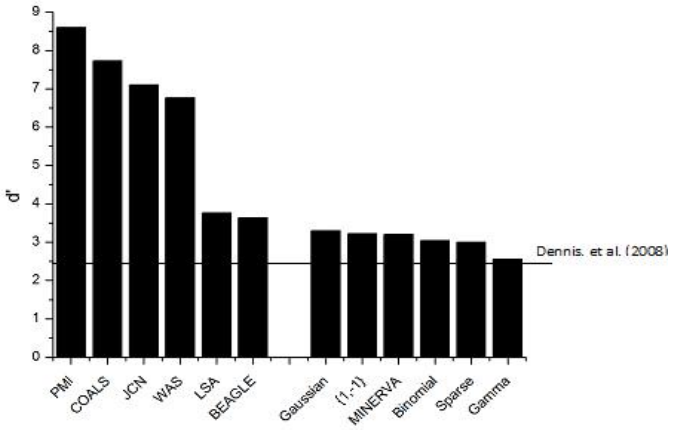

Figure 3. Levels of discriminability (d-prime) for SDT simulations; behavioral data from Dennis, et al. (2008).

Demonstration #2: MINERVA 2 and False Recognition

This demonstration was conducted in order to show that random representations provide an increase amount of freedom to fit data. The MINERVA 2 model of Hintzman (1986) has been used to successfully account for a variety of categorical false recognition effects (Arndt & Hirshman, 1998). Here, we simulate associative false recognition with the model, using both random and structured representations of semantics. Robinson and Roediger (1997) found that as the number of studied items that are related to a critical lure is increased, so is the probability of falsely recognizing that critical lure. The purpose of this demonstration is to compare the ease with which a simple process model like MINERVA is able to model this effect when using random representations versus when it is using representations that contain knowledge about the similarity structure of the actual words.

To construct MINERVA vectors that contain plausible semantic structure, we transformed the WAS representations from Steyvers et al. (2003). Typical applications of MINERVA use ternary vectors with a fairly low dimensionality. Hence, WAS vectors were collapsed from

400 to 20 dimensions by summing every 20 quadrants in the WAS vector into a single element in the reduced vector. This reduced vector was then transformed into a ternary vector with values of the set {-1, 0, 1}; the magnitude of the summed WAS values were recoded so that the highest third were assigned +1 (representing a high weighting on that feature), the middle third 0, and the lowest third -1. To ensure that the MINERA transformed vectors still reflected the semantic structure in the original WAS vectors, we computed the word-by-word cosines between vectors in both representations, and correlated the two matrices: The original vectors and their ternary transformed versions were highly correlated, $r = .67$, $p < .001$, indicating that the transformed vectors contain an arrangement of elements that reflects the semantic structure in the original WAS vectors. Using the false recognition lists from Stadtler, Roediger and McDermott, (1998) and Gallo and Roediger (2002), there was a high average similarity of the critical word's representation to the representations of the list items across the 52 word lists, $r = 0.35$, $p < .001$.

Random representations for critical words and their corresponding lists were created as in Arndt and Hirshman (1998), by using prototype and exemplar vectors. A prototype vector (representing the critical word) is first generated by randomly sampling elements from the set {1, 0, -1} with equal probability. Each item in the word list is then created by randomly perturbing elements in the prototype vector. This process requires a distortion parameter, which determines the probability of switching elements from the prototype vector when creating a list item vector. The distortion parameter determines how similar the list items are to the critical word. The important point is that both the semantic and random representations contain the exact same elements (same number of -1, 0, and 1s). The difference is that the elements are arranged independently for the random representations, whereas they are arranged to respect the inter-word similarity structure from WAS in the semantic version.

For MINERVA with a semantic representation, the results of Robinson and Roediger (1997) were modeled by randomly selecting 3 word lists, and adding 3, 6, or 9 items from one of the lists into a study list. Because the word lists in Robinson and Roediger were longer (they also used 12 and 15 associates), 27 words selected randomly from the Toronto word pool were added into the study list. To simulate this with MINERVA using random representation, 3, 6, or 9 exemplars were created for 3 random prototypes and added into the study list. Additionally, 27 random vectors were added into the study list to make the two simulations equivalent. Decisions are based on activation levels of a probe to the studied items (echo intensity: Hintzman, 1986), calculated by summing the similarity across all items in the study list.

For the MINERVA with semantic representations, there are two free parameters: 1) a criterion to make a new-old decision based on activation levels, and 2) a forgetting parameter which determines the probability of a non-zero element switching to zero during study. The simulation with random representations includes an additional distortion parameter (described above) to create the semantic structure. These parameters were fit to the data from Robinson & Roediger (1997) data using a Nelder-Mead simplex algorithm. The results of the simulation are displayed in Figure 4: the MINERVA model that utilizes random representations was able to reproduce the overall trend in the data. However, this was not the case with the MINERVA model that used semantic representations—this model tended to falsely recognize critical items over studied items, which is not the case with the human data. The random representation version of the model produced an excellent account of the data, $R^2 = 0.98$, $p < .001$. However, the version based on the true semantic similarity of the words used fit no better than chance, $R^2 = 0.05$, $p = .45$.

This simulation provides a simple demonstration of how a process model that has false representation assumptions may be incorrectly accepted as a plausible model. The only difference between the two models is in their representation structure—the process is identical. While the semantic version contains the "true" semantic structure for the exact words used in the experiment, the random version uses the distortion parameter to create the semantic structure that is most likely if this process account is correct. It is exclusively the incorrect inferred semantic structure that allows the process account to fit these data. If the correct representational structure were used, the process account would be rejected. The point is that random representations allow unnecessary freedom for the model to fit the data.

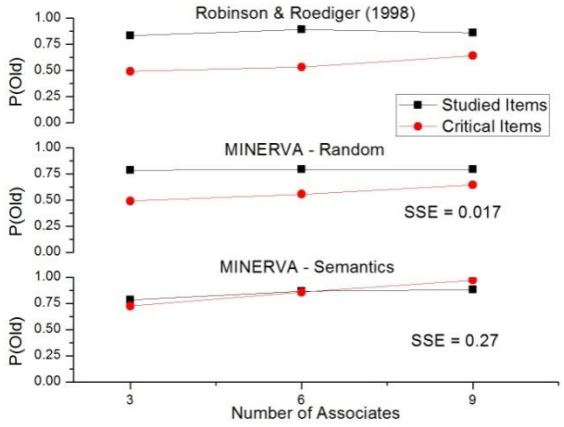

Figure 4. Results of false recognition simulation**.**

General Discussion

The use of randomly generated representations contains the assumption that semantic similarity is normally distributed over randomly selected pairs of words. This assumption was shown to be false across many different semantic metrics that have demonstrated success at accounting for human data. In experiments using words, two randomly selected words are likely to be relatively less similar (compared to the distribution of all possible pairs) than would be implied

using randomly generated representations for lexical semantics. Because similarity plays a central role in the processing mechanisms used by many memory models, the use of random representations may have consequences for conclusions drawn from simulations using these models.

As McClelland (2009) has noted, "…simplification is essential, but it comes at a cost, and real understanding depends in part on understanding the effects of simplification." (p. 18). The use of random representations in the development of cognitive models has been a necessary simplification for our understanding of cognitive processes. In doing so, researchers have made use of representations whose assumptions may not be entirely accurate, but through the use of this simplification modelers have made fundamental discoveries about how memory processes work. However without this assumption these results would not have been possible. It has only been within the last decade that researchers have had access to realistic representations of lexical semantics. The task for the future is to integrate semantic representations with processing models of memory for a fuller understanding of how they work together to produce observable behavior.

In accordance, recent models have begun to conduct this type of integration. For example, Monaco, Abbott, & Kahana (2007) have created a neural network model of the mirror effect of frequency, utilizing lexical semantic representations taken from the WAS of Steyvers, et al. (2004). Ideally, future models will combine a learning process that builds a representation through exposure to environmental information, which can then feed into a processing mechanism. For example, Johns and Jones (2009) have utilized representations built through a co-occurrence learning process to drive a processing model of both false recognition and false recall. These models suggest that it is no longer necessary to assume random representations for lexical semantics when modeling cognitive phenomena, but that item-specific semantic representations are now freely available and offer additional modeling constraints about the structure of semantic similarity that a process mechanism must operate on to produce behavior in a given task.

References

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations for a global matching perspective. *Journal of Memory and Language*, 39, 371-391.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition.*Psychological Review*, *108*, 452-478.

Dennis, S., Lee, M. D., & Kinnel, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language*, *59*, 361-376.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). Norms for Toronto word pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods and Instrumentation, 14*, 375-399.

Johns, B. T., & Jones, M. N. (2009). Simulating false recall as an integration of semantic search and recognition. *Proceedings of the 31st Annual Cognitive Science Society.*

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114*, 1-37.

Gallo, D.A., & Roediger, H.L. (2002). Variability among word lists in eliciting memory illusions: evidence for associative activation and monitoring. *Journal of Memory and Language, 47*, 469-497.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411-428.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent Semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review,104*, 211-240.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 3-23.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science, 1*, 11-38.

Monaco, J. D., Abbott, L. F., & Kahana, M. J. (2007). Lexico-semantic structure and the word-frequency effect in recognition memory. *Learning and Memory, 14*, 204–13.

Miller, G. A. (Ed.) (1990). WordNet: An on-line lexical database. *International Journal of Lexicography, 3*.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89*, 609-626.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. *Proceedings of the 17th Annual Conference of the Cognitive Science Society.*

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Recchia, G. L., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information to latent semantic analysis. *Behavior Research Methods,* 41, 657-663.

Robinson, K., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science*, 8, 389-393.

Rohde, D. L. T., Gonnerman, L., and Plaut, D. C. (submitted). An improved model of semantic similarity based on lexical co-occurrence. *Cognitive Science.*

Shiffrin, R.M. & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving Effectively from Memory. *Psychonomic Bulletin & Review, 4*, 145-166.

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

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004).Word Association Spaces for Predicting Semantic Similarity Effects in Episodic Memory. In A. Healy (Ed.), *Experimental Cognitive Psychology and its Applications.*