

Using Cognitive Models to Investigate the Temporal Dynamics of Semantic Memory Impairments in the Development of Alzheimer's Disease

Brendan T. Johns,¹ Vanessa Taler,² David B. Pisoni,¹ Martin R. Farlow,³ Ann Marie Hake,³ David A. Kareken,³ Frederick W. Unverzagt,⁴ Michael N. Jones¹

¹Department of Psychological and Brain Sciences, Indiana University

²Department of Psychology, University of Ottawa, and Bruyere Research Institute

³Dept. of Neurology, Indiana University School of Medicine, Indianapolis

⁴Dept. of Psychiatry, Indiana University School of Medicine, Indianapolis

Abstract

Mild cognitive impairment (MCI) is characterized by subjective and objective memory impairment in the absence of dementia. MCI is a strong predictor for the development of Alzheimer's disease (AD). A standard task used in the diagnosis of MCI is verbal fluency, where participants produce as many items from a specific category (e.g., animals) as possible. This task is typically analyzed by counting the number of items produced. However, analysis of the actual pattern of items produced can provide valuable additional information. Here we describe a computational model that uses multiple types of lexical information in conjunction with a standard memory search process. The model parameters were able to successfully differentiate changes in semantic memory processing for individuals who develop MCI, even when behavioral variables could not.

Keywords: Semantic memory; Alzheimer's disease; Verbal fluency; Computational modeling; Memory search

1. Introduction

Alzheimer's disease (AD) is a devastating neurodegenerative disorder that has a significant impact on memory function. Semantic memory is particularly impaired, and deficits are seen early in the disease course (Hodges & Patterson, 1995). In 1999, Petersen et al. described a syndrome referred to as mild cognitive impairment (MCI). This syndrome is characterized by subjective and objective memory impairment in the absence of dementia (Petersen, et al., 1999). The diagnosis of MCI is a strong predictor of the development of AD, with a conversion rate of 10-15% a year, compared with a conversion rate of 1-2% for the general population (e.g., Petersen, et al., 2001).

One of the most commonly used tasks to assess semantic memory in MCI and AD is category fluency, where the participant is asked to produce as many items from a specific semantic category as possible within a small amount of time. Fluency data are typically analyzed by counting the number of items produced: participants with AD and MCI tend to produce fewer items than healthy individuals (Taler & Phillips, 2008). However, this is not always the case, with some studies finding that significant differences only manifest themselves once individuals with MCI progress to AD (Lambon Ralph, et al., 2003). Thus,

the number of items produced is not always a stable predictor of MCI.

In addition, by examining only the number of items produced, little information can be inferred about the changes in the structure of semantic memory that accompany the development of AD and MCI. To examine changes in fluency output, Troyer et al. (1998) used a clustering-and-switching analysis, where the number of semantic clusters and the number of switches between clusters are counted with a hand-coding system. Troyer et al. found that individuals with MCI and AD produced less coherent clusters with fewer switches among the clusters.

Although the hand coding of clustering and switching has provided important insights about changes in the output of individuals who are developing or have AD, there are a number of issues with this type of analysis. Most importantly, clustering-and-switching analyses are subjective measures, raising issues of reliability and validity. This type of analysis is also expensive and time-consuming, given that a human rater (and multiple raters if reliability is an issue), must examine the fluency output of an individual subject. Given these problems, cognitive models may be promising tools for inferring the processes that produce fluency data, and they may also shed light on the changes that occur in semantic memory in AD and MCI. Cognitive models also afford automated coding to provide a standardized metric for comparison of semantic memory that is not dependent on a group of raters.

A number of models have been recently developed to model semantic fluency data, all based on vector space models of semantic representation. The optimal foraging model of Hills, Jones, & Todd (2012) proposes a clustering-and-switching mechanism, based on the BEAGLE model of semantic representation (Jones & Mewhort, 2007). A similar model is proposed by Abbot, Austerweil, & Griffiths (2012), who developed a random walk model over an associative semantic network.

The goal of the present paper is to use a model that integrates aspects of both the Hills et al. (2012) and Abbot et al. (2012) models in order to examine the longitudinal changes that occur in the development of MCI. Semantic fluency data were collected from participants at a memory disorders clinic; cognitively healthy participants are also assessed at this clinic. A cognitive model was used to assess the changes in searching behaviors that are seen with the

development of MCI. This model will serve as a theoretical tool to better understand changes that occur in memory searching behavior at the beginning stages of cognitive impairment, and also as a first step in the development of a practical tool that can be used to aid in the diagnosis of MCI, AD and other forms of dementia.

2. Computational Model

The goal of the model described here is to assess changes in the use of differential lexical information sources in individuals who develop MCI. Following the work of Hills et al. (2012), we model the path taken by participants through semantic space using Luce's (1959, 1977) choice axiom, a ubiquitous decision rule in cognitive psychology and economics. Rather than the full foraging model of Hills et al., we use a simplified multi-cue product integration model in which generation of a sequence of items is based on frequency and semantic similarity.

Multiple types of semantic information are used by the searching process. BEAGLE (Jones & Mewhort, 2007), which encodes both order and context information about the usage of a word, will form the lexical semantic representation of words. Additionally, rudimentary perceptual information will be attained by using the Generating Perceptual Representations (GPR) model of Johns & Jones (2012). The inclusion of this type of information will allow us to determine if inferred perceptual information allows for capture of additional variance in category fluency data. We now briefly describe these models.

2.1. Lexical Semantic Representation

Similar to other popular vector space models, BEAGLE (Jones & Mewhort, 2007) constructs lexical semantic representations by learning the structure of sentences in a large text corpus. Specifically, two types of information are learned: 1) context: the words that occur together with a given word across sentences, and 2) order: the respective position of other words to a given word in a sentence. Thus, the model learns both direct co-occurrence information, and also simple syntactic information about the use of a word in the language environment.

In BEAGLE, a word is represented with three vectors: 1) an environmental vector: a randomly generated vector that is assumed to represent the physical properties of a word, 2) a context vector, and 3) an order vector. Context information is learned by summing all of the environmental vectors that occur in the same sentence of a word.

In the original BEAGLE model, circular convolution of words in a sentence was used to learn order information. However, this technique is computationally expensive, and a technique using random permutations (RPs) was thus developed; this technique performs better due to its ability to scale up to larger corpora (Recchia, Jones, Sahlgren, & Kanerva, 2010). An RP is a random shuffling of an input vector (i.e. an environmental vector) to an output vector. Each position in a sentence is recursively permuted via the

unique RP function, and by permuting each word by the position that the word is in. These resulting RP vectors are summed into a composite order vector to represent the word's aggregate position relative to other words in the sentence.

To compute both context and order information, a 20 million sentence Wikipedia corpus was used. Environmental vectors were sparse ternary vectors, with a dimensionality of 10,000 and 6 non-zero items (sampled equally from 1 and -1).

In the fluency task described in section 3, the category used was that of animals (a standard category in the field). The corpus was preprocessed such that multiword animal names were concatenated into a single lemma in the corpus (e.g., "polar bear" was recoded as "polarbear"). Additionally, frequency information was obtained from this same corpus, and will be used in the model.

2.2. Perceptual Representation

A current development in the cognitive sciences is the realization of the importance of perceptual and grounded representations in language processing (Barsalou, 2008). Recently, a number of models have been developed that are able to construct inferences about the perceptual representations of words, based on the language environment and initial featural representations of a limited number of words. The model that will be used here is the GPR model of Johns & Jones (2012).

The GPR model uses global lexical similarity to construct perceptual representations about words that had no perceptual information in their lexical representations. The bases of the representation are the feature norms from McRae, et al. (2005). These norms only contain representations for around 500 words, but Johns & Jones (2012) used the global lexical similarity among words to generate perceptual representations. The result of this process is that all words have inferred perceptual representations, which were shown to be fairly accurate. Even though these representations are imperfect, this process does allow for simple perceptual information to be included in the memory searching process described below.

GPR representations were generated for each animal produced in the fluency experiment. The same Wikipedia corpus on which BEAGLE was trained was used to generate these representations. The corpus was reduced to 250,000 documents, due to the computational costs associated with the perceptual inference model.

2.3. Processing Mechanism

As described previously, the searching mechanism used here is based on the Luce choice rule (Luce, 1959, 1977), a well-known decision mechanism. The axiom defines how humans probabilistically select an item from possible alternatives (in our case, the word produced from all the words that could have been produced). Given a set of stimulus similarities, Luce's axiom states that the

probability of responding to stimulus S_i with response R_j is defined as:

$$P(R_j|S_i) = \frac{\beta_j^{\lambda_0} S(i, j)^{\lambda_1}}{\sum_{k=1}^n \beta_k^{\lambda_0} S(i, k)^{\lambda_1}} \quad (1)$$

where β_j is the response bias for item j , and $S(i, j)$ is the similarity between item i and j . The parameters λ_0 and λ_1 control the relative contributions of base rate (frequency) and similarity in producing the response (both are positive real values). In our simulations, the “stimulus” is the previous word produced, and the “response” is the word that is about to be produced (from the set of alternatives that could be produced). The set of alternatives that will be considered is all of the category members that subjects in the experiments produced in the fluency task that is being modeled.

In this model, there will be multiple types of similarity information used, including context, order, and perceptual similarity (built with the models described above). Similarity is computed as vector cosine, which is a length-normalized dot product. We use a model comparison paradigm to test whether each information source adds an increase in fit to the human data. The parameters of the model will control how much attention is allocated to a particular information source. Given that we are using three different similarity types, the full Luce rule may be expressed as:

$$P(w_i | w_{i-1}) = \frac{\beta_i^{\lambda_0} \prod_{j=1}^3 S_j(w_{i-1}, w_i)^{\lambda_j}}{\sum_{k \in \text{Animals}} \beta_k^{\lambda_0} \prod_{j=1}^3 S_j(w_{i-1}, w_k)^{\lambda_j}} \quad (2)$$

Where w_i is the current word, w_{i-1} is the previous word, β_i is the log-frequency of the current word, and s_1 , s_2 , and s_3 are context, order, and perceptual similarity, respectively. Each of the $\lambda_0 \dots \lambda_3$ parameters control the importance of their respective information sources. Additionally, similarity is transformed with exponential scaling (e.g. $exp^{s_{sim}}$), because this transformation yields a better fit to most similarity-based behaviors (Shepard, 1987).

2.4. Parameter Fitting and Model Testing

For each participant’s sequence of items produced, we determine the most likely set of parameters that would have generated the observed data if the model were correct. This will be done longitudinally across multiple tests. We compare a variety of nested models in their goodness-of-fit to the fluency data, each representing a potential cognitive process that may have generated the data.

All possible models will be tested. With four parameters, there are a total of 12 different models. Parameters were fit for each participant under each of the above models using

maximum likelihood estimation (Myung, 2003). Specifically, a grid-search algorithm was used to find the optimal set of parameters to maximize the log-likelihood that the model generated the data. All parameter values between 0 and 30, in steps of 1.0 were tested. The value of 30 was selected because no participant exceeded this value, for either the semantic similarity or the frequency parameter. This method allows the best-fitting parameters of a single model for a particular individual to be determined.

Models were compared using the Akaike information criterion (AIC; Akaike, 1974), a standard and reliable method to compare models’ ability to quantitatively fit the human data. The AIC compares the quantitative fit of a model to human data (based on the above log-likelihood), intrinsically penalizing models as a function of the number of free parameters. Models with the lowest AIC value are preferred, and this value will be used to select among the different proposed models. By differentiating between models we investigate the most salient information sources that are used in memory search to produce fluency data.

2.5. Discussion

The model described here provides a simple mechanism with which to examine the use of different information sources in category fluency. Examining parameter changes across multiple linguistic information sources, and multiple verbal fluency sessions (at different time points), affords insight into the cognitive changes that accompany dementia.

3. Verbal Fluency Study

3.1. Participants

Two participant groups were included in the present study. First, we identified all healthy older adults in the Indiana Alzheimer Disease Center database who received an eventual diagnosis of amnesic mild cognitive impairment (aMCI), and for whom neuropsychological data were available at least two years prior to diagnosis ($n=13$). Healthy older controls were then matched individually to the aMCI patients for age (± 5 years) and education (± 3 years).

Participants had no neurological or psychiatric history other than aMCI. The diagnosis of aMCI was made using a consensus conference format composed of psychiatrists, neurologists, and neuropsychologists and was based on a review of the clinical assessment material. Criteria were consistent with Petersen (2004) as follows: (a) informant-reported or physician-detected decline in cognition or memory, or (b) psychometric test scores below the approximately the 7th percentile of same-age peers, and (c) no significant impairment in activities of daily living. Average number of assessments and number of years followed, demographic characteristics and baseline neuropsychological performance are provided in Table 1.

Table 1
Descriptive statistics of groups

	Pre-MCI Participants	Control Subjects
N	13	13
Sex	10 men, 3 women	10 men, 3 women
Age at final assessment	77.69 ± 6.80	75.77 ± 6.39
Education	15.23 ± 3.35	14.31 ± 3.12
# of assessments	3.31 ± 2.02	3.00 ± 1.22
Years of follow-up	4.85 ± 1.82	4.46 ± 2.18
Estimated full-scale IQ	112.46 ± 9.68	113.31 ± 5.30
Mini-mental state examination (/30)	28.54 ± 1.20	28.85 ± 1.21

3.2. Procedure

Participants completed a verbal fluency task as part of a neuropsychological battery completed in the memory clinic. The battery lasted approximately 2-3 hours. The verbal fluency task proceeded as follows. Participants were told that they would receive a category and be asked to produce as many items as possible within that category in one minute. They were then told that the category was “animals”, and the experimenter recorded manually each item that the participant produced.

3.3. Results

Cognitively healthy participants produced slightly more items on average ($M = 18.42$; $SD = 2.06$) than did pre-MCI participants ($M = 17.45$; $SD = 3.28$), but the difference was not statistically significant. However, mean total items produced ignores differences across time in the number of items outputted. To compute this, the slope of the number of items produced across testing sessions was calculated. Healthy controls had a mean slope of -0.12 ($SD = 0.29$), while pre-MCI patients had a mean slope of -0.12 ($SD = 0.32$). Again, while the slopes were on average steeper for participants in the pre-MCI group, the difference in slope across the two groups was not statistically significant.

For this sample of pre-MCI and cognitively healthy control participants, the number of items produced was not a significant indication of the development of cognitive impairments. Neither was the change in the number of items produced (assessed by taking the slope of the number of items produced across trials).

4. Computational Analysis

4.1. Model Fits

The AIC values were computed for all 12 possible models. However, due to space limitations only the best 2- and 3-parameter models will be displayed. Figure 1 displays the result of the fits for pre-MCI (top panel) and control participants (bottom panel). This figure shows that both the control and pre-MCI participants have the same trends in

terms of the pattern of best-fitting models. It also shows that all information sources provide a better fit over a random model (where going to any word in the category is equally probable). The best one-parameter model was order information, with frequency information being the second most important. The best two-parameter model is a combination of order similarity and frequency. The best three-parameter model is a combination of order similarity, perceptual similarity, and frequency. The complete model actually performs slightly worse than the three-parameter model, due to the penalty that AIC provides for additional parameters. This suggests that sentence context information does not account for any unique variance above and beyond the other three parameters. Thus, the three-parameter model is the most parsimonious account, and the parameters of this model will be analyzed in section 4.2.

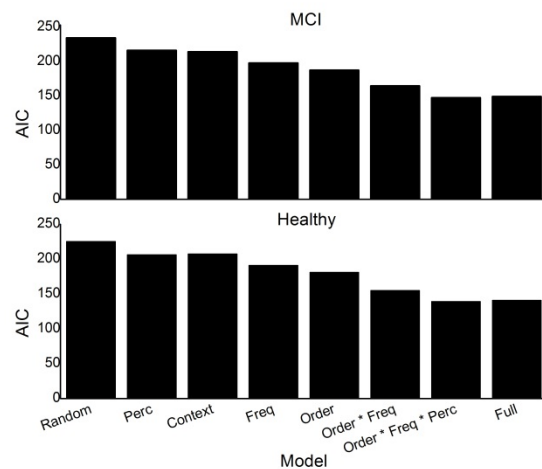


Figure 1. Model fits for both MCI and healthy individuals.

The model fits suggest that in terms of semantics, sentence order information is more predictive of searching patterns than sentence context information. This suggests that subjects are more likely to go to a word that was used in a similar linguistic manner as the preceding item, rather than simply a word that occurred in the same sentence (i.e., paradigmatic similarity is favored in this task over syntagmatic similarity). Additionally, even though the type of perceptual similarity that we are using is rather low in resolution, it still adds power to the model. This finding has not been reported in previous literature, and suggests that verbal fluency is not entirely based on linguistic information, but rather that other types of information are also used. As would be expected, frequency is also used heavily in the searching process: high frequency words tend to be selected more than low frequency ones. However, the most important aspect of this analysis is how parameters change across time for both control participants and pre-MCI participants.

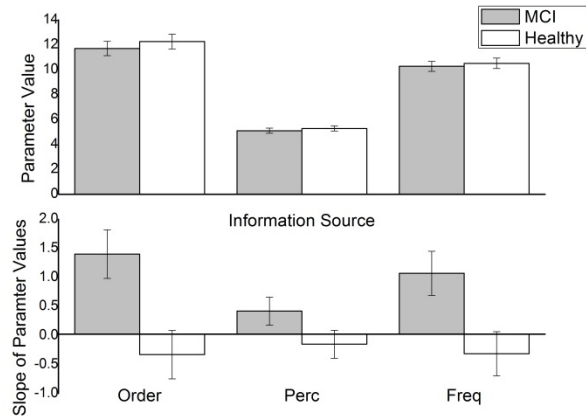


Figure 2. Average parameter values (top panel) and slope of parameters across trials (bottom panel) for pre-MCI and healthy individuals. Error bars represent standard error.

4.2. Parameter Analysis

The first analysis aimed to determine if there is any difference in mean parameter values for control and pre-MCI participants. The top panel of Figure 1 displays the mean parameter values for the three parameter types and two participant groups. This figure shows that control participants have a slightly higher value on all three parameters relative to pre-MCI participants. To test whether these differences were significant, all trials across both control and pre-MCI participants were added into a multivariate ANOVA with subject type as a between subjects factor. By including all of the trials in the test, a stronger test of parameter differences can be done.

There was no significant difference found for the order parameter [$F(1,124)=0.469$, *ns*], frequency parameter [$F(1,124)=0.544$, *ns*], or the perceptual parameter [$F(1,124)=0.684$, *ns*]. Hence, there was no reliable difference the mean parameter values between pre-MCI and control participants.

However, as mentioned in the behavioral result section, the mean parameter does not take into account how the parameters are changing across time. In order to assess this, the slope of the parameters across sessions was calculated, and the bottom panel of Figure 2 displays the results of this analysis. This figure shows that for all 3 parameters the slope of the parameter is positive for pre-MCI participants, but negative for healthy controls. This difference is statistically significant for the order similarity parameter [$F(1,24)=8.577$, $p<0.01$] and the frequency parameter [$F(1,24)=6.671$, $p<0.05$], but not significant for the perceptual similarity parameter [$F(1,24)=2.797$, $p=0.1$].

The above finding suggests that over time, participants who will go on to be diagnosed with MCI begin to generate higher frequency words that are closer together in semantic space (reflected by the increase in the order parameter). Even though they are still producing the same total number

of items as healthy controls, the pattern of items produced is more likely to be guided by high frequency cues and they are more likely to produce items from closely connected regions of semantic memory (somewhat analogous to the findings of Troyer et al., 1998). Cognitively healthy control participants' longitudinal data are more likely to be generated by a model with no changes in parameter values over time (for healthy controls the average slope was not statistically different from zero). Even though there is no detectable difference in the behavioral variables (e.g., number of items produced) between MCI and healthy controls, the changes in model parameters most likely to produce the individual pattern of items produced was sensitive to group differences. That is, although the sequences *robin-worm-snake* and *robin-sparrow-chicken* both sum to three items, the parameters that determine the different paths taken through memory may differ greatly, and capture additional variance that is sensitive to early group differences.

Given that MCI patients do not benefit from practice on verbal fluency tasks (Cooper, et al., 2004), the parameter changes seen are indicative of underlying changes that are occurring in the semantic memory system. Additionally, this analysis suggests that perceptual information is more stable with memory degradation than is purely linguistic information, as the slopes for the perception parameter were not significantly different across groups. Ober and Shenaut (1999) found that there was no deficit for individuals with AD in conceptual processing of animals. Given that conceptual features are a part of the McRae norms, it is possible that this type of information is less impaired in individuals with MCI and AD. However, the use of superior perceptual representations of words will have to be used in order to determine if this is a true effect.

5. Discussion

This paper describes a model-based analysis that evaluates changes in semantic memory occurring across time prior to an individual's being diagnosed with MCI, which is often a precursor to the development of Alzheimer's disease. This model uses a standard decision mechanism, a generalization of the Luce choice rule, together with multiple sources of information about words learned from vector space models, to model semantic fluency. The best fitting model was found to be a cue integration model that combines sentence order information from BEAGLE (Jones & Mewhort, 2007), perceptual information from the GPR model (Johns & Jones, 2012), and frequency. Sentence context information was also tested, but did not provide a superior fit over the simpler three-parameter model.

The behavioral data consisted of 13 participants who went on to develop MCI and 13 who remained cognitively healthy. They were assessed with verbal fluency tasks in annual checkups. Neither the total number of items produced, nor the slope of changes in number of items produced across time was able to differentiate those participants who went on to develop MCI and those who did

not. In addition, the average parameter values of the model also could not differentiate the two groups. However, the changes in parameter values across time showed a significant interaction. Specifically, the order and frequency parameters tended to increase across time for the MCI participants, but were stable for the cognitively healthy controls. There was a difference in the perceptual parameter, but this was not large enough to reach significance, suggesting that this type of information is not impacted as strongly as purely linguistic information. However, given the limitations in the assessment of perceptual similarity that the model used, more research is required to determine the importance of perceptual and grounded knowledge in memory search.

This analysis demonstrates that a simple memory search model that utilizes semantic information about words is capable of assessing the temporal changes involved in the development of MCI, even when behavioral data cannot. Thus, cognitive modeling can play an important role in understanding complex clinical data, and has the potential to provide tools for the early detection of cognitive disorders.

The results reported here have important implications for the early diagnosis of AD, and for identification of older adults at risk of developing cognitive impairment. While measures of total output in category fluency, as typically used in clinical practice, were unable to identify older adults who would go on to be diagnosed with MCI, the modeling approach we used was able to identify quantitative differences between pre-MCI participants and cognitively healthy controls. We suggest that more sophisticated analysis of standardly available clinical data may prove to be a fruitful approach in the diagnosis of cognitive decline. Alterations in searches within semantic space are observed as early as seven years prior to diagnosis with MCI, indicating that subtle cognitive changes may occur many years prior to clinical memory dysfunction.

6. Acknowledgements

This research was supported by a grant from the Indiana Clinical Translational Sciences Initiative (NIH-TR000006) to MNJ and VT, by NSF BCS-1056744 to MNJ, by a Young Investigator Award from the Alzheimer Society of Canada to VT, and by NIA P30 AG10133 to the Indiana Alzheimer's Disease Center. BTJ was supported by a postgraduate scholarship from NSERC.

7. References

Abbott, J., Austerweil, J., & Griffiths, T. (2012). Human memory search as a random walk in a semantic network. *Advances in Neural Information Processing Systems*, 25, 3050-3058.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.

Cooper, D. B., et al. (2004). Category fluency in mild cognitive impairment: Reduced effect of practice in test-retest conditions. *Alzheimer's Disease and Associated Disorders*, 18, 120-122.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, 119, 431-440.

Hodges, J. R., & Patterson, K. (1995). Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia*, 33, 441-459.

Johns, B. T., & Jones, M. N. (2012). Perceptual inference from global lexical similarity. *Topics in Cognitive Science*, 4, 103-120.

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

Lambon Ralph, M. A., et al. (2003). Homogeneity and heterogeneity in mild cognitive impairment and Alzheimer's disease: A cross-section and longitudinal study of 55 cases. *Brain*, 126, 2350-2362.

Luce, D. R. (1959). *Individual Choice Behavior*. New York: John Wiley & Sons.

Luce, D. R. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, 15, 215-233.

McRae, K., et al. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37, 547-559.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47, 90-100.

Ober, B. A., & Shenaut, G. K. (1999). Well-organized conceptual domains in Alzheimer's disease. *Journal of the International Neuropsychological Association*, 5, 676-684.

Peterson, R. C., et al. (1999). Mild cognitive impairment: Clinical characterizations and outcome. *Archives of Neurology*, 56, 303-308.

Petersen, R. C., et al. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, 58, 1985-1992.

Recchia, G. L., et al. (2010). Encoding sequential information in vector space models of semantics: Comparing holographic reduced representation and random permutation. *Proceedings of the 32nd Annual Cognitive Science Society*. Austin TX: CSS.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317-1323.

Taler, V., & Phillips, N. A. (2008). Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. *Journal of Clinical and Experimental Neuropsychology*, 30, 501-556.

Troyer, A. K., et al. (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of the International Neuropsychological Association*, 4, 137-143.