Cognitive Modeling as an Interface between Brain and Behavior: Measuring the Semantic Decline in Mild Cognitive Impairment

Brendan T. Johns,[1] Vanessa Taler,[2] David B. Pisoni,[3] Martin R. Farlow,[4] Ann Marie Hake,[4] David A. Kareken,[4] Frederick W. Unverzagt,[5] Michael N. Jones[3]

[1]Department of Communicative Disorders and Sciences, University at Buffalo
[2]School of Psychology, University of Ottawa, and Bruyère Research Institute
[3]Department of Psychological and Brain Sciences, Indiana University
[4]Dept. of Neurology, Indiana University School of Medicine, Indianapolis
[5]Dept. of Psychiatry, Indiana University School of Medicine, Indianapolis

Word Count: 5,461

Correspondence
Dr. Brendan Johns
Dept. of Communicative Disorders and Sciences
122 Cary Hall
University at Buffalo
Buffalo, NY, 14214

Email: btjohns@buffalo.edu
Phone: (716) 829-2797
Fax: (716) 829-3979

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, and may represent an early stage in the disease course in many cases. 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. Verbal fluency performance is typically analyzed by counting the number of items produced. However, analysis of the semantic path of the items produced can provide valuable additional information. We introduce a cognitive model that uses multiple types of lexical information in conjunction with a standard memory search process. The model used a semantic representation derived from a standard semantic space model in conjunction with a memory searching mechanism derived from the Luce choice rule (Luce, 1977). The model was able to detect differences in the memory searching process of patients who were developing MCI, suggesting that the formal analysis of verbal fluency data is a promising avenue to examine the underlying changes occurring in the development of cognitive impairment.

Keywords: cognitive modeling, mild cognitive impairment, semantic memory, language modeling, verbal fluency

Cognitive modeling is a fundamental tool used to understand the processes that underlie behavior, and has become a standard technique in the cognitive sciences (Busemeyer & Diederich, 2010; Lewandowsky & Farrell, 2010). However, it is typically the case that cognitive modeling is used for theoretical purposes, as a method to derive formal accounts of the mind, and not necessarily as a tool to aid or solve an applied problems (there are exceptions of course, such as automated essay grading using distributional models of semantics; e.g. Foltz, Laham, & Landauer, 1999). To demonstrate the usefulness and potential of cognitive models it is important to take models that have been developed to explain a cognitive or behavioral phenomenon, and apply them to unresolved problems in similar domains. That is, cognitive models should be both theories of cognition and cognitive tools that can be used to analyze behavior.

One area that has received a great deal of attention in cognitive modeling is the verbal fluency task (Abbott, Austerweil, & Griffiths, 2015; Hills, Todd, & Jones, 2012; Taler, Johns, Sheppard, Young, & Jones, 2013). In this task, the participant is required to produce as many items as possible from a specific category (e.g., animals) within a given time period (typically one minute). This task has received attention because it provides an interesting insight into the patterns of memory search and retrieval, which other standard tasks do not provide.

The verbal fluency task is also widely used in the neuropsychological literature to assess semantic memory functioning (for a review, see Taler & Philips, 2008). In particular, it is a focus of tests that are designed to assess the development of dementia (particularly Alzheimer's disease) and mild cognitive impairment. Mild cognitive impairment (MCI; Petersen et al., 1999) is a relatively recent clinical diagnostic category, which 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 Alzheimer's disease (AD) and other dementias, 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). Hence, MCI is often seen as an early marker in a continuum of decline that culminates in dementia, making detection of MCI a crucial goal to target early intervention.

While declines in semantic memory functioning are not necessary for a diagnosis of MCI (Petersen et al., 1999; Albert et al., 2011), many studies have found deficits in semantic memory for patients diagnosed with MCI, including verbal fluency and comprehension measures (Adlam, Bozeat, Arnold, Watson, & Hodges, 2006), object and person naming (e.g. Ahmed, Arnold, Thompson, Graham, & Hodges, 2008; Joubert, et al., 2010), conceptual processing of famous people and events (Joubert, Felician, Barbeau, Didic, Poncet, & Ceccaldi, 2008), and longitudinal changes in semantic memory performance (Mickes, et al., 2007), among others (for a review, see Taler & Phillips, 2008). The exact nature of the semantic impairment in MCI remains unclear. There is evidence for both content and process accounts; that is, people with MCI may experience both semantic control deficits and broad taxonomic loss of semantic knowledge (Reilly, Peelle, Antonucci, & Grossman, 2011; Taler, Voronchikhina, Gorfine, & Lukasik, 2014).

In terms of the verbal fluency task, it is well established that individuals with AD produce fewer items than healthy controls (see Taler & Phillips, 2008). However, the pattern is less clear in MCI: while some studies find that people with MCI produce significantly fewer items than controls, other studies have found that significant differences only emerge once individuals with MCI progress to AD (e.g., Lambon Ralph, et al., 2003).

The standard approach in analyzing this task is to simply count of the total number of items produced in the fluency task. This is a coarse measure that ignores potentially diagnostic information contained in the path taken through semantic memory. Consider two short sequences of animals: *dog, cat, mouse* and *dog, wolf, deer*. A count of the sequences collapses both to the same number, and ignores the semantic variance inherent in the path. The semantic cohesion of the path taken is thought to be a particularly salient component of the early stages of memory degradation.

In order to examine the semantic pathway that a subject takes during a fluency experiment, Troyer, et al. (1998) developed a method of analysis in which the number of semantic clusters and the number of switches between clusters are counted. Clusters are identified using a hand-coding system. It has been found that patients with AD and MCI produce less coherent clusters with fewer switches among the clusters (Troyer, et al., 1998; Murphy, Rich, & Troyer, 2006). While many important insights have been obtained using this method, there are some drawbacks, mainly that coding is time-consuming and requires multiple human raters for reliability.

With the advent of computational cognitive models of verbal fluency (e.g. Hills, et al., 2012), it is now possible to have automated scoring measures of verbal fluency performance. That is, it is possible to take the theories that have been developed to explain this type of behavior and use it in a more applied setting as a cognitive tool to better understand the type of underlying changes that are occurring in a clinical population.

The underlying operation of models such as Hills, et al. (2012) and Taler, et al. (2013) is the use of *semantic space models* (SSMs) of lexical semantic memory (Jones, Willits, & Dennis, 2015) to form a basis of word meanings. More specifically, Hills, et al. (2012) proposed a

memory foraging model that uses a mechanism similar to Troyer et al.'s (1998) clustering-and-switching methodology, but is fully automated and operates on a memory representation learned by a semantic space model trained on a large linguistic corpus. The model can make predictions for any sequence of items produced in a fluency task, and can infer the most likely weighting of cognitive variables used by the individual to generate the observed pattern of items. A similar method has been used to assess the semantic memory system of bilinguals (Taler, et al., 2013), by analyzing changes in the use of different linguistic information types across fluency tasks. Additionally, a recent paper by Hills, Mata, Wilke, and Samanez-Larkin (2013) also used this type of model to examine age-related changes in memory searching strategies, demonstrating the power of this approach to analyze different participant types.

The purpose of the current work is to use a simple memory search model, similar to Hills et al. (2012), 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 were 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. The goal of this research is to provide insights into the underlying changes in the semantic memory system that are occurring in people who are developing MCI. We use a formal analysis, which can shed light on the particular cognitive variables that are changing over time as cognitive impairment develops. This will provide an increased level understanding about the changes occurring in semantic memory in patients with MCI, and also offer a new cognitive tool to analyze verbal fluency data.

Modeling Fluency Performance

The primary function of semantic space models is to take very large pieces of language (e.g. encyclopedias, collections of books, newspapers, etc…) and use these materials to construct representations of word meanings (ideally for every word in a language). These representations have provided an additional tool to researchers: the ability to analyze behavior at the item level. This has led to a variety of important insights, including into semantic priming (Jones, Kintsch, & Mewhort, 2006; Hare, et al., 2009), bilingual language switching (Taler, et al., 2013), false memory (Johns, Jones, & Mewhort, 2012), sentence processing (Johns & Jones, 2015), among many others. SSMs have begun to be used to examine deficits in clinical populations (e.g. Hoffman, Rogers, & Lambon Ralph, 2011), which is coherent with the goals of the current model. The reason why this ability to analyze item-level behavior is important in examining verbal fluency performance, is that it allows for the path of the production to be analyzed, which allows for a greater amount of information to be attained about a complex behavior.

The model used here aims to assess changes in the use of differential lexical information sources in people who go on to develop MCI. This will be done by collecting a number of lexical variables, namely a variety of similarity values and word frequency, and using these variables in a simple and canonical memory search model. The use of these different lexical variables in the searching mechanism will be assessed in a parameter fitting and model testing framework, which will allow for a determination of the most important variables used in verbal fluency, and also how these variables are changing in the development of cognitive impairment.

Specifically, following early work by Romney, Brewer, and Batchelder (1993), we model the path taken by participants through semantic space using Luce's (1977) choice axiom, a ubiquitous decision rule in cognitive psychology and economics. Rather than basing our estimate of memory space on subjective ratings, as Romney et al. did, we use multiple sources of

information learned by semantic models. Specifically, we estimate linguistic similarity between items using the BEAGLE model (Jones & Mewhort, 2007), which learns both contextual and role information about a word's usage in language, and we estimate perceptual similarity between items using the perceptual inference model of Johns and Jones (2012).

Conceptual Representation

Similar to other popular semantic space models, BEAGLE (Jones & Mewhort, 2007) constructs lexical semantic representations by observing statistical redundancies in a large text corpus. This model works at the sentence level and records the usage of a word by learning two types of statistical information: *context* (the words that co-occur with a given word in language, e.g., *cat-mouse*), and *order* (the shared temporal roles of words with respect to other words, e.g., both *cat* and *panther* pounce on prey). This information is stored in a large, distributed vector, somewhat akin to the hidden layer in a neural network. Context information is similar to pure co-occurrence information, and marks the probability of two words occurring together, while order information is more akin to simple syntactic information, akin to how a word is used in language. The model was trained to learn both types of information on a 20-million sentence corpus taken from Wikipedia. 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"). Frequency information was obtained from the same corpus.

An additional information source integrated into the model is perceptual similarity of the animal exemplars, estimated by a model proposed by Johns and Jones (2012) that constructs inferred perceptual representations for words based on the feature norms of McRae, et al. (2005). The original McRae et al. norms contain subject-generated feature vectors for only 133 animals,

but the perceptual inference model uses lexical similarity among words to infer the perceptual feature vector for animals that were not originally normed with a high degree of accuracy.

The information that is taken from these models are word-word similarity values. Similarity was assessed with a vector cosine, which is a normalized dot product and gives a value between -1 and 1. These values were normalized into a range of 0 to 1 in order to be able to be used in probability calculations in the memory search mechanism. The technical details of the BEAGLE and the GPR models can be found in Appendix A and B, respectively.

Processing Mechanism

Our model of fluency makes decisions about sequences of animals to produce based on the Luce choice axiom (Luce, 1977). The axiom defines how humans probabilistically select an item from possible alternatives, such as selecting the word *dog* from the set of {*cat*, *dog, horse*, …, *zebra*} in a fluency task. The model will use two types of information: the similarity value between the previous word and a set of possible words, and a bias factor (a base rate production value, assumed to be word frequency; normalized between 0 and 1 by dividing all frequency values by the most frequent word). Formally, Luce's axiom states that the probability of responding to stimulus S_i (previous word produced) with response R_j (word about to be produced) 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). The set of alternatives is all of the category members that the subjects in the experiments produced.

As an example of how this memory search model works, consider that a participant has three animals in their lexicon: *dog*, *wolf*, and *horse*. Assume these words have a normalized frequency value of 0.8, 0.4, and 0.6 respectively. The first word is produced by sampling randomly based on the normalized frequency, as there is no similarity value possible, and this led to *dog* being produced. The word *dog* had a 44.4% chance of being produced first:

$$\frac{0.8}{0.8 + 0.4 + 0.6} = 0.444$$

To produce the next word, both the similarity between dog-wolf and dog-horse and the bias factors are used in the probability calculations. To calculate this, assume that the semantic similarity between *dog-wolf* is 0.8 and *dog-horse* is 0.3. The words *wolf* and *horse* would then have the following probabilities, as derived from equation (1):

$$\Pr(wolf|dog) = \frac{0.4 * 0.8}{(0.4 * 0.8) + (0.6 * 0.3)} = 0.64$$

$$\Pr(horse|dog) = \frac{0.6 * 0.3}{(0.4 * 0.8) + (0.6 * 0.3)} = 0.36$$

Thus, wolf and horse would have a 64% and 36% chance of being produced, respectively, under these conditions.

Our model uses multiple information sources when making a decision (frequency, context, order, and perceptual similarity). Similarity between two words is computed as the cosine between their length-normalized vectors. The parameters of the model control how much attention is allocated to a particular information source when producing a word. Given that we are using three different similarity types, the full Luce rule may be expressed as:

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

where w_i is the current word, w_{i-1} is the previous word, β_i is the normalized log-frequency of the current word, and s_1, s_2, and s_3 are context, order, and perceptual similarity, respectively. Each of the $\lambda_0 \ldots \lambda_3$ parameters control the importance of their respective information sources. It has been found that transforming the values with an exponential function provide superior fit of this type of model to verbal fluency data (Taler, et al., 2013), and so this has also been done here. This leads to the canonical form of the model being described with the following equation:

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

Another advantage of using this type of transformation is that it is now the case that as you increase the value of a parameter it causes a corresponding increase in the importance of that information source in the memory searching operation, allowing for easier visualization of the contributions of different information sources.

Parameter Fitting and Model Evaluation

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. To understand this process, consider the example of *dog-wolf-horse* above. Using the same similarity and frequency values, and the exponential transformation used in equation (3), *wolf* has a 57.5% probability of being produced after *dog*, while *horse* has a 42.5% chance, when λ is set at 1 for both. However, if the λ for semantic similarity is set at 4 while the λ for frequency remains at 1, the probability values switch to 86% for *wolf* and just 14% for *horse*, due to the increased import of semantic similarity in the searching operation. However, if these parameter values are

switched to emphasize the importance of frequency, these probabilities change to 42.5% for *dog* and 57.5% for *horse*, reflecting a shift towards sampling from base rates. By optimally fitting the parameters to an individual's verbal fluency output in this manner, a look into the type of information that is being used by the person is found.

This evaluation is performed longitudinally for each subject during multiple annual follow-up 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. As stated, the models used 4 different types of information in memory search: 1) order similarity, 2) context similarity, 3) perceptual similarity, and 4) frequency. Each information type has its own parameter within the searching mechanism. With four parameters, there are a total of 12 different models that have to be tested. 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.

Models were compared using the Akaike information criterion (AIC; Akaike, 1974), a standard method to compare models' ability to quantitatively fit human data (Shiffrin, Lee, Kim, & Wagenmakers, 2008). The AIC compares the quantitative fit of a model to human data (based on 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. Parameters were fit to each individual across the different testing sessions.

The model described above provides a simple mechanism with which to examine the use of differential information sources, derived from the environmental structure of language, in

category fluency. To demonstrate how different models will predict different retrieval strategies, Figure 1 displays the predicted output from a limited set of animals {*cat*, *lion*, *robin*, *sparrow*, *chicken*} for three different models (frequency, frequency * order, and frequency * order * perceptual). For the frequency only model, the output is simply determined by relative frequency. The two-parameter model generates its output using lexical similarity and frequency: for instance, *cat* and *robin* occur in more similar lexical contexts than *cat* and *lion*. However, when perceptual information is included in the production mechanism, *cat* now cues *lion* due to greater feature overlap of their percepts. Given that the order and context similarity values likely account for similar types of variance it is questionable whether these two sources will both provide significant power in accounting for verbal fluency performance, but it is predicted that both frequency and perceptual similarity should account for large amounts of variance. By estimating the usage of various nested information sources, application of this model provides insight into the memory retrieval strategies used in the task, and also the dynamic changes that are occurring in the development of MCI.

Method

All data was were collected prior to diagnosis of an aging disorder (no participants had any pre-existing neurological or psychiatric conditions), so there were no impairments in the patient's ability to make informed consent. Additionally, all data were analyzed anonymously, so linking an individual patient with their data was not possible.

Participants

The overall goal of our study was to formally examine the longitudinal changes in semantic memory that occur as MCI develops. In order to accomplish this, we identified every cognitively healthy older adult in the Indiana Alzheimer Disease Center database who received an eventual diagnosis of amnestic mild cognitive impairment (aMCI), and for whom neuropsychological data were available at least two years prior to diagnosis. The two-year cutoff was based on the annual assessment done in the memory clinic, which ensured the availability of at least two data points for each participant prior to diagnosis. We then identified healthy older controls who were best-matched on an individual basis to the eventual aMCI patients. MCI and control participants were matched for age (±5 years), education (±3 years), and sex. The final sample comprised 13 people with MCI, and 13 matched controls.

Even though our final sample had only an n=26, this is a very well controlled subset. A larger sample could have been used, but at the cost of reduced power (error variance would be introduced to the effect) and the introduction of confounds in causal variables (see Wagenmakers, et al., 2014 for a discussion on issues of power in the behavioral literature). The increased control in our dataset increases the likelihood that any differences between the groups are due to the underlying cognitive change occurring in the MCI group, rather than to an unidentified confound. Hence, even though the size of the pre-experimental sample was significantly reduced by our selection criteria, this was countered by ensuring that our sample would be reflective of changes occurring in pre-MCI patients. Because our aim is to assess cognitive changes occurring across time in memory search processes in people undergoing cognitive change, by reducing noise in our sample we hope to increase the probability that the model described above will find an informative difference between the two groups.

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 approximately the 7th percentile of age- and education-matched 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.

Procedure

Participants completed a verbal fluency task as part of a neuropsychological battery completed in the memory clinic. The full battery lasted approximately 2-3 hours. For the verbal fluency task, participants were instructed 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.

Results

Behavioral Analyses

Mean number of items produced by cognitively healthy participants (M = 18.42; s = 2.06) was statistically equal to the number produced by pre-MCI patients (M = 17.42; s = 3.28), $F(1,24)$= 0.80, *ns.* However, analyses using mean total items ignore potentially important changes across assessments in the number of items produced. We therefore computed the slope of the number of

items produced across testing sessions for each participant. Again, the mean slope of items produced across testing sessions was statistically equal between healthy controls (M = -0.01, s = 0.29) and pre-MCI patients (M = -0.29, s = 0.32), $F(1,24) = 0.75$, *ns.* For this sample of pre-MCI and cognitively healthy control participants, neither number of items produced nor the decrease in number of items produced across assessments was a significant indication of the development of MCI.

There are a number of possible reasons for the lack of behavioral differences seen in the data reported here. One possibility is that the repeated testing that was used could have led to practice effects, leading to no differences in number of items produced. Another possibility is that our sample had a high average IQ (112.46), which may have protected this sample from deficits in verbal fluency production performance. However, the goal of this study was to assess differences in the semantic pattern of items produced by our subjects, not simply the number of items produced. We suggest that, while behavioral data is very important in assessment and diagnosis, cognitive modeling can provide additional insights into the cognitive changes that occur as cognitive impairment develops.

Cognitive Model Fits

Figure 2 displays the AIC fits for pre-MCI (top panel) and control participants (bottom panel) for the best 2- and 3-parameter models. As the figure shows, both the control and pre-MCI participants have the same trends in terms of the pattern of best-fitting models. For both groups, all information sources provide a better fit over a random model (where transitions among words are equiprobable). The best model for each group was the three-parameter model that integrated order similarity, perceptual similarity, and frequency. A Bayes factor conversion of AIC (Berger

& Pericchi, 1996) indicated that the three-parameter model was 1,259 times more likely than the two-parameter model to have generated the data, and 49 times more likely than the full model. Thus, our analysis of the parameters will focus on the three-parameter model.

The model fits demonstrate that in terms of semantic similarity based on linguistic experience, order information (shared role) is more predictive of searching patterns than simple context information. 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 suggests that verbal fluency is not entirely based on linguistic information, but grounded perceptual information may play a role as well. As was expected, frequency was also an important search cue to predict transitions.

Parameter Analyses

As a first examination into the item-level patterns of production for the control and patient groups that the modeling analysis can be compared against, a category analysis using Troyer categories was done. To do this, the extended Troyer categories from Hills, Jones, & Todd (2012) was used to assess both the number of categories sampled by each participant, and also the slope of the changes in number of categories produced. The average number of categories produced for the control participants was 12.35, while the pre-MCI participants produced 13.01 categories per fluency session, a non-significant effect [$F(1,25)=1.528$, *n.s.*]. The slope of the change in number of categories produced was also assessed, in order to examine the temporal changes associated with the development of MCI. Control participants had an average slope of 0.0637, while pre-MCI participants had an average slope of -0.169, also a non-significant effect [$F(1, 25)=0.689$, *n.s.*]. This indicates that whatever changes are occurring in the patient group, an

analysis examining the overall categories that are being produced in not sensitive enough to detect them.

For the cognitive model, our first analysis of interest explores whether the best-fitting parameters differ between control and pre-MCI participants. The top panel of Figure 3 displays the mean parameter values for the three parameter types between groups. Although control participants have a slightly higher value on all three parameters relative to pre-MCI participants, a multivariate ANOVA indicated no significant differences between groups in any of the parameters (order parameter: $F(1,124) = 0.469$, *ns*; frequency parameter: $F(1,124) = 0.544$, *ns*; perception parameter: $F(1,124) = 0.684$, *ns*), nor any significant interactions. Echoing the behavioral analysis, the average cognitive parameters most likely to have generated the fluency data were insensitive to the difference between pre-MCI and control participants.

As discussed in the behavioral result section above, however, the mean parameter does not take into account how the parameters change across assessments. To examine change, the slope of the parameters across assessments was computed; the results are displayed in the bottom panel of Figure 3. For all three parameters, the slope of the parameter was positive for pre-MCI participants, but negative for healthy controls. This difference was statistically significant for the order similarity parameter [$F(1,25) = 8.702$ $p < 0.01$; $\eta^2 = 0.266$] and the frequency parameter [$F(1,25) = 6.803$, $p < 0.05$; $\eta^2 = 0.221$], but not for the perceptual similarity parameter [$F(1,24) = 2.87$, $p > 0.1$; $\eta^2 = 0.107$]. To ensure that this effect was not simply due to the final measurement (when patients were diagnosed with having MCI) a subsequent analysis was conducted with the final session removed from the calculation of slope. All effects remained the same, with differences in order [$F(1,25) = 7.379$, $p < 0.01$; $\eta^2 = 0.235$] and frequency [$F(1,25) = 6.056$, $p < 0.05$; $\eta^2 = 0.201$] remaining significant, but no difference in the use of perceptual information

[$F(1,25) = 2.188$, $p > 0.1$; $\eta^2 = 0.084$]. This demonstrates that the model can detect differences over time in the memory searching patterns of cognitively healthy older adults from those who will go one to be diagnosed with MCI.

General Discussion

This article describes a model-based analysis of the changes in semantic memory occurring across time prior to a person 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 word frequency. Context information from BEAGLE was also tested, but did not provide a superior fit over the simpler three-parameter model, suggesting that it was too redundant with order information to provide a unique contribution to the memory searching process.

The behavioral data came from 13 participants who went on to develop MCI and 13 who remained cognitively healthy. They completed verbal fluency tasks in annual assessments. Neither the total number of items produced, nor the slope of changes in number of items produced over time, was able to differentiate those participants who went on to develop MCI from 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 differed significantly between the groups. Specifically, the order and frequency parameters increased over time for the pre-MCI participants, but were stable for the cognitively healthy controls. That

is, even though the two groups produce the same total number of items, the pattern of items produced by pre-MCI participants is more likely to be guided by high frequency cues that are from more closely connected regions of semantic memory. This same pattern has previously been found to separate older subjects from the demented and very old (Morais, Neth, & Hills, 2013), which points to converging evidence for this result. 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.

There was a difference in the perceptual parameter, but this difference was not large enough to reach statistical significance, suggesting that this type of information is not impacted as strongly as lexical-semantic information. However, given the limitations in the assessment of perceptual similarity within the model, more research is required to determine the importance of perceptual and grounded knowledge in memory search. It is worth noting, however, that conceptual processing (which is an aspect of the McRae et al. norms) is less impaired in MCI and AD than other types of semantic knowledge (Ober & Shenaut, 1999), suggesting that this type of information may be minimally impacted early in the disease course, compared with other types of language-based information.

Although there were no detectable differences between the groups in the behavioral measures (e.g., number of items produced), the changes in model parameters across assessments were able to differentiate the groups. For example, although the sequences *robin-worm-snake* and *robin-sparrow-chicken* both include three items, the parameters that determine the different paths taken through memory differ greatly, and the modeling approach described here captures additional variance that is sensitive to early group differences. The inclusion of cognitive

modeling in the analysis of verbal fluency data may thus provide valuable additional information about semantic function that is not necessarily available directly from the behavioral data. It also demonstrates that cognitive models are a powerful tool in a behavioral scientist's toolbox that allows for more information to be extracted from data, which surface level examinations are unable to provide.

One important distinction to make about the results of this study is whether the parameter differences found between the healthy controls and pre-MCI patients reflect executive function changes or degradation of lexical representations in the development of cognitive impairment. Previously, we have used the same model reported here to examine language switching in young bilinguals (see Taler, et al., 2013), a task that is thought to heavily involve executive functioning. It was found that the model is sensitive to behavioral changes that are reflective of executive functioning changes. Specifically, it was found that when you force bilinguals to repeatedly switch between languages they use frequency information at a much greater rate than pairwise semantic similarity. That is, there seems to be an active switch in strategy that is used in memory search. So, within the context of past results, it is possible that the deficits results from a change in executive functioning performance, rather than say a degradation of lexical representations. However, based on the results reported in this paper it is not possible to put support to either proposition, and to separate the two positions will require additional empirical and modeling work.

This analysis demonstrates that a simple memory search model that utilizes semantic information about words is capable of assessing temporal changes in semantic function prior to the development of MCI. The model was able to find alterations in the path that the different groups were taking through semantic space, indicating that subtle cognitive changes may occur

many years prior to clinical memory dysfunction. This suggests that cognitive modeling can play an important role in understanding complex clinical data, and is a promising tool to explore the underlying changes occurring in cognitive impairment.

References

Adlam, A. R., Bozeat, S., Arnold, R., Watson, P., & Hodges, J. R. (2006). Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease. *Cortex*, 42, 675-684.

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

Ahmed, S., Arnold, R., Thompson, S. A., Graham, K. S., & Hodges, J. R. (2008). Naming of objects, faces and buildings in mild cognitive impairment. *Cortex*, 44, 746-752.

Albert, M. S., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia, 7, 270-279.

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

Berger, J. O, & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91, 109-22.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Sage.

Clark, D. G., Kapur, P., Geldmacher, D. S., Brockington, J. C., Harrell, L., Deramus, T. P., Lokken, K., Nicholas, A. P., & Marson, D. C. (2014). Latent information in fluency lists predicts functional decline in persons at risk for Alzheimer disease. *Cortex*, 55, 202-218.

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

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

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

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

Foltz, P. W., Laham, D., & Landauer, T. K. (1999). Automated essay scoring: Applications to educational technology. *Proceedings of EdMedia*, 99, 40-64.

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

Johns, B. T., & Jones, M. N. (2015). Generating structure from experience: A retrieval-based model of sentence processing. *Canadian Journal of Experimental Psychology*, 69, 233-251

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

Jones, M. N., Willits, J. A., & Dennis, S. (2014). Models of semantic memory. In J. R. Busemeyer & J. T. Townsend (Eds.) *Oxford Handbook of Mathematical and Computational Psychology*.

Joubert, S., Felician, O., Barbeau, E. J., Didic, M., Poncet, M., & Ceccaldi, M. (2008). Patterns of semantic memory impairment in mild cognitive impairment. *Behavioural Neurology*, 19, 35-40.

Joubert, S., Brambati, S. M., Ansado, J., Barbeau, E. J., Felician, O., Didic, M., Lacombe, J., Goldstein, R., Chayer, C., Kergoat, M. J. (2010). The cognitive and neural expression of

semantic memory impairment in mild cognitive impairment and early Alzheimer's disease. *Neuropsychologia*, 48, 978-988.

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.

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. Sage.

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

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers*, 37, 547-559.

Mickes, L., Wixted, J. T., Fennema-Notestine, C., Galasko, D., Bondi, M. W., Thal, L. J., & Salmon, D. P. (2007). Progressive impairment on neuropsychological tasks in a longitudinal study of preclinical Alzheimer's disease. *Neuropsychology*, 21, 696-705.

Murphy, K. J., Rich, J. B., & Troyer, A. K. (2006). Verbal fluency patterns in amnestic mild cognitive impairment are characteristic of Alzheimer's type dementia. *Journal of the International Neuropsychological Society*, 12, 570–574.

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.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science, 4,* 28-34.

Sahlgren, M., Holst, A., & Kanerva, P. (2008). Permutations as a means to encode order in word space. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, p. 1300-1305.

Salmon, D. P., Butters, N., & Chan, A. S. (1999). The deterioration of semantic memory in Alzheimer's disease. *Canadian Journal of Experimental Psychology*, 53, 108-117.

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.

Taler, V., Johns, B. T., Young, K., Sheppard, C., & Jones, M. N. (2013). A computational analysis of semantic structure in bilingual verbal fluency performance. *Journal of Memory and Language*, 69, 607-618.

Troyer, A. K., Moscovitch, M., Winocur, G., Leach, L., & Freedman, M. (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of the International Neuropsychological Association*, 4, 137-143.

Wagenmakers, E.-J., Verhagen, A. J., Ly, A., Bakker, M., Lee, M. D., Matzke, D., Rouder, J. N., & Morey, R. D. (2014). A power fallacy. *Behavior Research Methods*.

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.

Table 1

Descriptive statistics of groups

	Pre-MCI Participants	Control Participants
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

Figure Captions

Figure 1. Examples of item sequences in a fluency task predicted by progressively higher parameter models. The same set of four items is used to demonstrate how the most likely sequence generated would change as additional information sources are added.

Figure 2. Quantitative model fit (Akaike Information Criterion) of various nested versions of the fluency production model for pre-MCI participants (top panel) and healthy controls (bottom panel). A smaller AIC indicates a better fit to the human data.

Figure 3. Average parameter values (top panel) and slope of parameters across trials (bottom panel) for pre-MCI and healthy individuals using the most likely three-parameter model. Error bars represent standard error.

Figure 1.

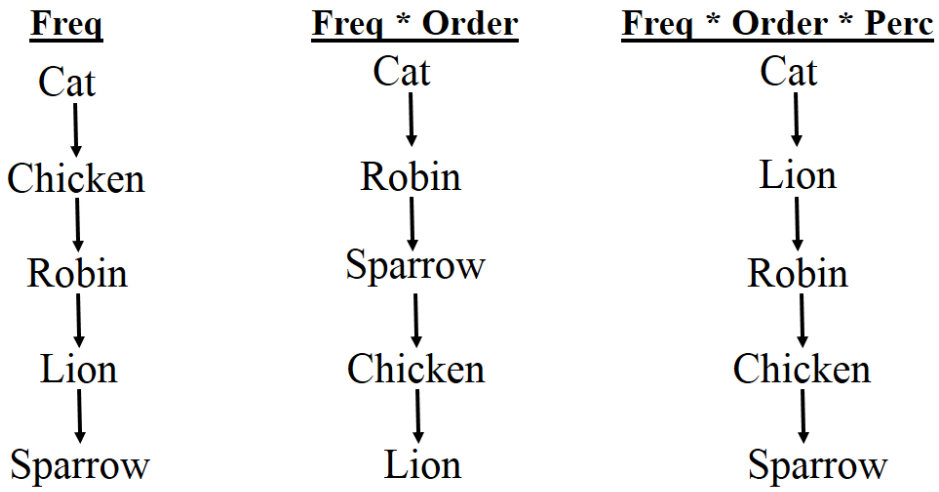

Freq
Cat
Chicken
Robin
Lion
Sparrow
Freq * Order
Cat
Robin
Sparrow
Chicken
Lion
Freq * Order * Perc
Cat
Lion
Robin
Chicken
Sparrow

Figure 2.

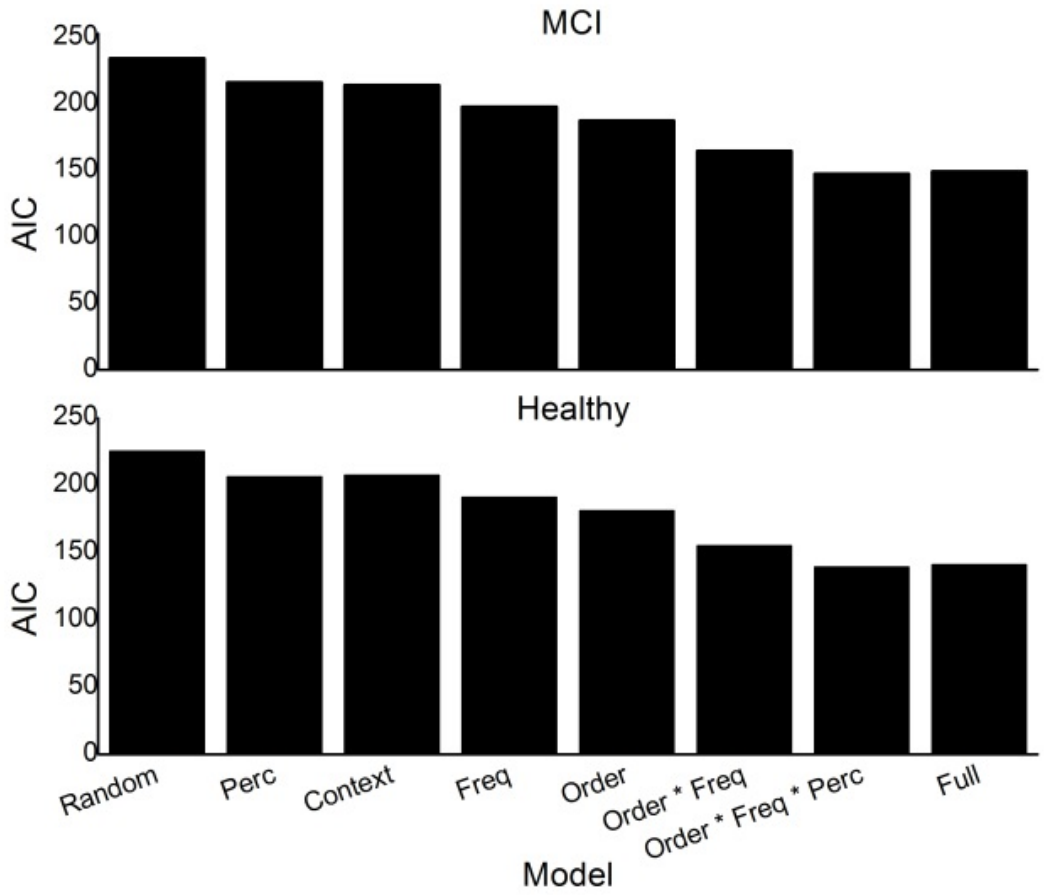
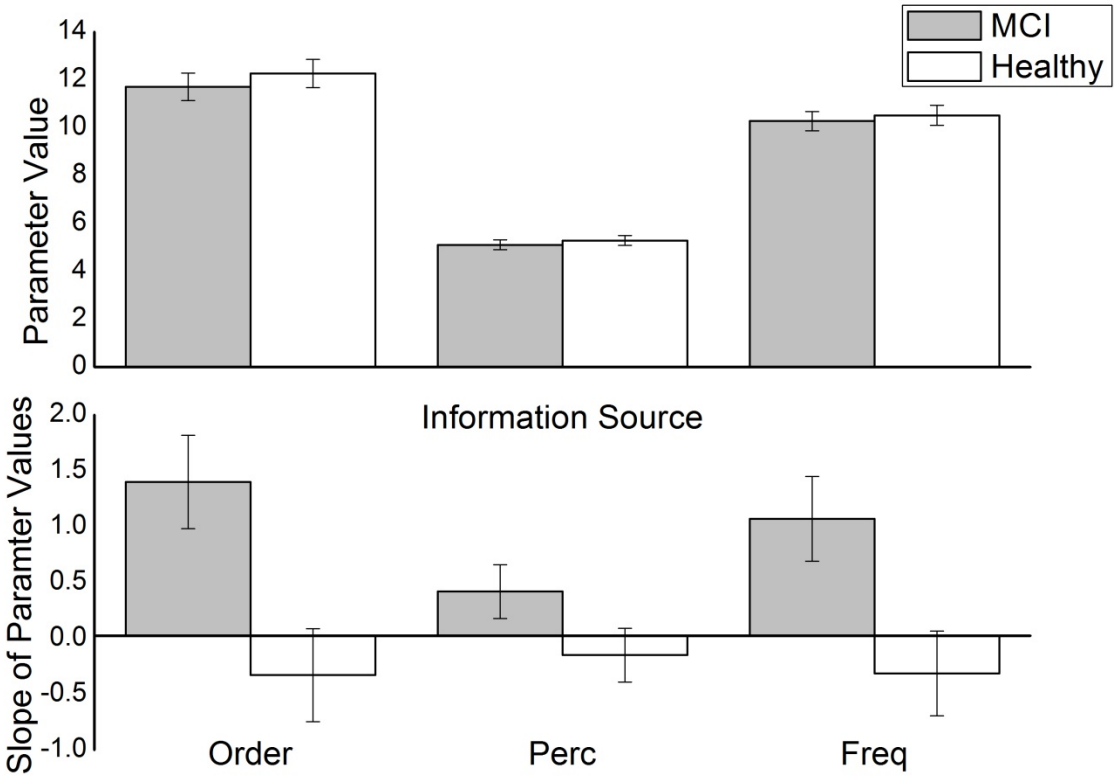

MCI
AIC
250
200
150
100
50
0
Healthy
AIC
250
200
150
100
50
0
Random
Perc
Context
Freq
Order
Order * Freq
Order * Freq * Perc
Full
Model

Figure 3.


MCI
Healthy
Parameter Value
14
12
10
8
6
4
2
0
Information Source
Slope of Paramter Values
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
Order
Perc
Freq

Appendix A

BEAGLE-RP Model

Originally, the BEAGLE model used circular convolution of words in a sentence was used to learn order information (Jones & Mewhort, 2007). However, this technique is computationally expensive, so a technique using random permutations (RPs) was thus developed (Sahlgren, et al., 2007). This technique has been shown to perform better, as it is capable of scaling up to larger corpora (Recchia, Jones, Sahlgren, & Kanerva, 2010), so it is used here. As in the original model, words are initially represented with a static environmental vector, which is assumed to represent perceptual properties of a word. However, instead of the Gaussian representation that the original model utilizes, BEAGLE-RP uses sparse ternary vectors (non-zero locations sampled equally from 1 and -1). The environmental vectors used in the current analysis had a dimensionality of 10,000 with 6 non-zero items. The environmental vectors are used to build both context and order vectors, which are dynamic vectors that change across time. Context vectors represent pure co-occurrence information across a corpus. These vectors are computed for a given word in a sentence, c_i, by summing the environmental vectors for the other *n-1* words in the sentence:

$$C_i = \sum_{j=1}^{n} e_j \quad where\ i \neq j$$

Where n goes through each word in the sentence.

Order vectors are assumed to represent rudimentary syntactic information, by recording the position of word's in a sentence relative to the word being updated, and here random permutations of environmental vectors are used to learn this information type. A random permutation is simply a random shuffling of an input vector into an output vector. Each location within a sentence is given its own permutation, which does not change across training, allowing

for a simple method of encoding word order. For any given word's order vector, o_i, this is updated by permuting the word's that surround the word o. For a word o, in position i, this process is described with the following equation:

$$o_i = \sum_{j=1}^{n} RP^{j-i}\, e_j$$

Where RP^x is a specific RP for a given location. Negative RPs are simply the inverse of their positive counterparts, and are used to encode the respective position of words within the sentence (i.e. if a word occurs before or after the word under question). Order vectors are updated for each word in the sentence.

Appendix B

The Generating Perceptual Representations (GPR) Model

The technique we employ here to generate perceptual similarity values are attained from the Generating Perceptual Representations (GPR) model described in Johns & Jones (2012). The GPR uses global lexical similarity to construct perceptual representations about words that had no perceptual information in their lexical representations, based on the processing of a classic model of memory (Hintzman, 1986). The bases of the representation are the feature norms from McRae, Cree, Seidenberg, & McNorgan (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 cross-validated with reasonable precision.

At the beginning of the process, each word in the lexicon has a pure co-occurrence representation, which is represented with a large Word x Document matrix, where words are contained in rows and documents in columns. A word gets a value of 1.0 if it occurs in a document, and a value of 0.0 if it does not. The featural representation for the 500 words from the McRae norms are concatenated onto their respective lexical representations. For words that were not contained in the norms, their perceptual representations were initially left empty.

In the first step of the model, each representation in memory with a zero perceptual vector has an estimated perceptual vector constructed based on its weighted similarity to lexical entries that have non-zero perceptual vectors:

$$Perc_j = \sum_{i=1}^{M} T_i * S\left(T_i, T_j\right)^{\lambda}$$

Where M represents the size of the lexicon, T represents the lexical trace for a word, S is similarity function (here, vector cosine), and λ is a similarity weighting parameter. In the second step of the model, the process from step 1 is iterated, but inference for each word is made from global similarity to all lexical entries (as they all now contain an inferred perceptual vector). Hence, representations in step 1 are inferred from a limited amount of data (only words that have been "perceived" by the model). In step 2, representations for each word are inferred from the full lexicon—aggregate linguistic and perceptual information inferred from step 1. Johns & Jones (2012) demonstrated that this model is able to make accurate perceptual inferences, so here these inferred perceptual representations are used to construct rudimentary perceptual similarity values into a memory searching process.