UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

A Biologically Constrained Model of Semantic Memory Search

Permalink

https://escholarship.org/uc/item/1n71300b

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors

Kajic, Ivana
Gosmann, Jan
Komer, Brent
et al.

Publication Date

2017

Peer reviewed

A Biologically Constrained Model of Semantic Memory Search

Ivana Kajić, Jan Gosmann, Brent Komer, Ryan W. Orr, Terrence C. Stewart, Chris Eliasmith
Centre for Theoretical Neuroscience, University of Waterloo
Waterloo, ON, Canada N2L 3G1
{i2kajic, jgosmann, bjkomer, rworr, tcstewar, celiasmith}@uwaterloo.ca

Abstract

The semantic fluency task has been used to understand the effects of semantic relationships on human memory search. A variety of computational models have been proposed that explain human behavioral data, yet it remains unclear how millions of spiking neurons work in unison to realize the cognitive processes involved in memory search. In this paper, we present a biologically constrained neural network model that performs the task in a fashion similar to humans. The model reproduces experimentally observed response timing effects, as well as similarity trends within and across semantic categories derived from responses. Three different sources of the association data have been tested by embedding associations in neural connections, with free association norms providing the best match.

Keywords: semantic memory; associations; semantic search; spiking neural network; neural engineering framework

Introduction

The semantic memory system plays an important role in a variety of cognitive functions. It is essential for language comprehension and understanding, and has been referred to as a *mental thesaurus*, storing knowledge about words, their meaning and relationships among them (Tulving, 1983).

The advent of neuroimaging techniques and observations from brain lesion studies have allowed more specific localization of the brain regions and networks responsible for semantic representation and processing (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Quiroga, 2012). In particular, the medial temporal lobe and portions of anterior lobes have been identified as essential to the function of semantic memory. Purely computational semantic network models have successfully explained behavioral data (Collins & Quillian, 1969; Collins & Loftus, 1975) and have been purported to reveal principles guiding language formation and organization (Steyvers & Tenenbaum, 2005). Yet, they have been severely limited in their ability to account for the neural realization of such processes. Our understanding of *how* networks of millions of neurons perform the computations that underly semantic processing is still extremely limited.

We propose a network of simulated spiking neurons that is able to perform the semantic fluency task in a manner similar to humans. While providing a good match with behavioral data, the model also proposes specific neural mechanisms that may be involved in semantic processes. The components of the model are discussed in terms of functionally and neurologically plausible counterparts found in the human brain.

Search in the Semantic Space

The semantic fluency task has been used to understand how humans search memory when asked to retrieve items semantically related to a given cue (Thurstone, 1938; Bousfield & Sedgewick, 1944). In a typical trial, a person is instructed to generate members of a category within a given time limit. One common version of the task requires an individual to list all animals they can think of within a fixed timespan of one or more minutes. Response analysis shows they tend to be grouped into clusters corresponding to subcategories (Troyer, Moscovitch, & Winocur, 1997), such as *pets* or *farm animals*. For example, responses might start with the animals an individual is most familiar with, such as *cat*, *dog*, *rabbit* and then continue with a list of farm animals such as *cow*, *chicken* and *turkey*.

To explain the clustering trend observed in the responses, Hills, Jones, and Todd (2012) suggested that individuals generate responses according to the optimal foraging policy (Charnov, 1976). Animals use such a strategy when searching for food in natural environments: after resources in one area have been depleted, animals continue their search for food in a new patch. In the context of the semantic fluency task, an individual listing animals in a specific sub-category would stop listing animals from that category after being unable to generate new items at a certain rate. Search behavior suggestive of optimal foraging has been reproduced with several different representations and algorithms, including a random walk on a semantic network constructed from free association norms (Abbott, Austerweil, & Griffiths, 2015). Jones, Hills, and Todd (2015) attribute the simplicity of this particular algorithm to the association norms being a direct result of an experimental design that is very similar to the semantic fluency task. They argue that the fundamental memory retrieval processes and representations are obscured by the data underlying the model and the behaviors that are being explained. However, association data from sources other than association norms, like data learned from natural language, have successfully been used to reproduce human response patterns with random walks (Nematzadeh, Miscevic, & Stevenson, 2016).

Here, we take a first step towards explaining how the memory retrieval processes and representations described above can be realized by a biologically constrained neural network. The proposed model performs the search based on associative weights encoded within connections between neurons, resembling aspects of a random walk while still conforming to constraints of neural computation. The noise resulting from spiking neurons and the diversity in neuron parameter values lead to the response variability. We show that the search patterns observed in the model responses are consistent with the optimal foraging theory and match human behavioral data.
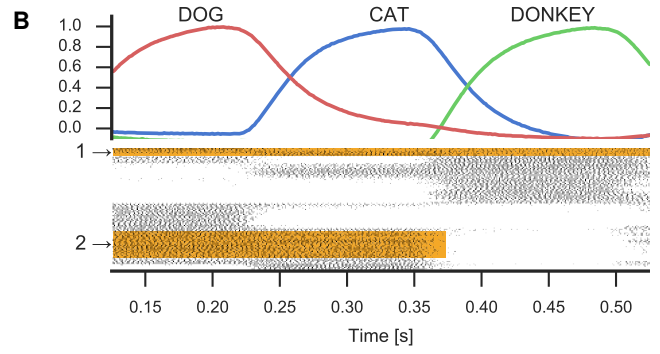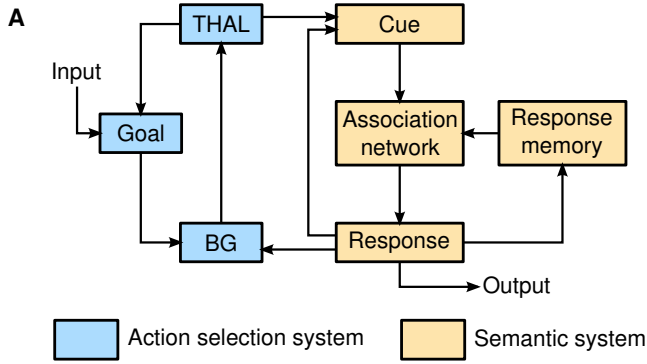
Figure 1: **A**: Architecture of the neural network model performing the semantic fluency task. Each box represents a population of spiking neurons. **B**: Neuronal spiking activity in the model recorded from the population *cue*. Some neurons are actively spiking when representing words *dog*, *cat* and *donkey* (highlighted area 1), while others only spike when representing words *dog* and *cat* (highlighted area 2). The similarity between these spikes and the ideal spike pattern for each word is shown above.

Biologically Constrained Representation

Brain imaging studies provide evidence in support of semantic representations distributed across networks of neurons in various brain regions (Huth et al., 2016; Rissman & Wagner, 2012). While many neurons jointly contribute to representations, single neurons can still exhibit preference for certain input stimuli. For example, neurons in the medial temporal lobe show selective responses for higher-level semantic concepts such as places or people (Quiroga, 2012).

Consistent with the notion of a distributed representation, we employ vector-based representations that can be implemented in a network of spiking neurons by means of the Neural Engineering Framework (NEF; Eliasmith & Anderson, 2003). In the NEF, connection weights between neurons can be analytically computed such that the neural network approximates a desired function.

Given an *n*-dimensional vector representing a preferred stimulus e and some time-varying input x, the activity of a single neuron a_i can be expressed as

$$a_i = G_i \left[\alpha_i e_i^\top x + J_i^{\text{bias}} \right] \qquad (1)$$

where G represents a spiking neuron model, in this case the Leaky Integrate-and-Fire (LIF) model. The parameter α scales the input and converts the unit of the variable (x) to units of current, and J^{bias} represents background currents.

As a result, if a neuron is driven by an input x that is similar to its preferred direction e, the dot product $e^\top x$ is larger ($e^\top$ is a transposed vector e). For a LIF neuron, this translates to a higher input current that drives the neuron to produce a more rapid series of spikes that is transmitted to another neuron. In biological systems, spikes are transmitted across synaptic connections and transformed to post-synaptic current at the site of a receiving neuron. It is important to note that the inputs to the neuron do not have to be characterized as scalar values, as Equation 1 holds for vector inputs.

We can recover the value represented by populations of neurons by filtering spike trains with a filter $h(t)$ and scaling with decoding weights d_i:

$$\hat{x} = \sum_i a_i * [d_i h]. \qquad (2)$$

The linear filter $h(t) = \tau_{\text{syn}}^{-1} \exp(-t/\tau_{\text{syn}})$ models the post-synaptic current. The symbol $*$ denotes convolution, an operation that places such filter at every position where a spike occurs, and sums the result. The decoding weights d_i can be analytically computed by a least-squares minimization of the error term $E = \|x - \hat{x}\|$.

To perform a computation, these decoding weights are coupled with the encoding weights e of the receiving neurons. This gives observable connection weights between two neural populations. Specifically, the connections between neurons in the pre-synaptic population a_i and the post-synaptic population b_j are computed as $w_{ji} = \alpha_j e_j^\top d_i$. The group of receiving neurons can also represent a transformed value $f(x)$, where f can be a non-linear function. The same optimization method can be used in this case to compute alternative decoding weights d_i^f to estimate the function.

Representing Words and Associations

In our model, the vectors x in Equation 1 are 256-dimensional unit vectors that represent animal words. The vectors are generated randomly such that similarity between any two vectors is generally less than 0.1. This ensures almost orthogonal vectors, with some overlap in representation, meaning the same neurons will be involved in the representation of different words. An example of this representational overlap can be seen in the spike raster plot in Figure 1B, where some neurons fire for all words and some only for a subset. The NEF methods allow us to decode the spiking activity in terms of the words being represented by the neurons with Equation 2 as shown in the upper part of Figure 1B.

Associative relationships between words are represented as linear transformations implemented in the connections be-

Table 1: Utility calculations for different goals and the corresponding actions.

	Goal	Utility calculation	Action
1.	Start	$goal \cdot \textbf{start}$	Set *cue* to **animal**, set *goal* to **think**
2.	Think	$goal \cdot \textbf{think} + \text{response\_magnitude} - 1$	Copy *response* to *cue*, add *response* to *responses*, set *goal* to **think**
3.	Default	0.4	Set *cue* to **animal**, set *goal* to **think**

tween two groups of neurons. Word vectors are collected row-wise into a single matrix V, and associations between pairs of words are encoded into a matrix A such that A_{ij} is the association strength from word i to word j. We can then express a new matrix $\tilde{A} = V^\top A^\top V$ to implement a transformation that multiplies the vector represented by the first group of neurons by the matrix $\tilde{A}$ and transmits the result to the second group. This operation results in a weighted linear combination of vectors that represents words associated with the word represented in the first group of neurons. This method of representing associations is embedded in a large recurrent network to perform the semantic fluency task.

Association Matrices To construct three different association matrices A, we use three different sources of associative data: Free Association Norms (FAN; Nelson, McEvoy, & Schreiber, 2004), BEAGLE (Jones & Mewhort, 2007) and Google Ngrams (Michel et al., 2011).

The FAN data set has been derived empirically in a free association experiment, where individuals were asked to generate the first word which comes to their mind for given a cue. The data was normed over all participants to yield asymmetric association strengths for over 5,000 words. The Ngram data set contains co-occurrences of sequences of n words extracted from the Google Books Ngram Viewer dataset (Version 2 from July 2012, Michel et al., 2011). This dataset provides occurrence frequencies of n-grams across over 5 million books published up to 2008. We use occurrences of bi-grams to construct an asymmetric association matrix. The BEAGLE dataset has been trained on a 400M-word Wikipedia corpus, yielding unique vector representations for each word. In this data set, similarity between pairs of vectors is computed as cosine similarity, providing a symmetric measure of association strength. We use pre-computed similarities between pairs of animal word-vectors as in Hills et al. (2012).

We take human responses as a reference for the set of animal words and consider only words that are present in all datasets, amounting to 157 animals. The FAN data set contains the smallest vocabulary and is the most restrictive set.

Proposed Neural Network Model

Using the NEF implemented in the Nengo simulation environment (Bekolay et al., 2014), we constructed a model consisting of approximately 62,000 LIF neurons organized in functional subgroups performing the semantic fluency task.[1]

The architecture in Figure 1A shows how networks of neurons are organized and connected to perform the task. The model can be divided into two components: the semantic system and the action selection system. In terms of their biological correlates, the semantic system can be mapped to areas of the medial temporal cortex, and the action selection system to the basal ganglia and the thalamus. The action selection system maintains two possible phases: initializing the task and performing the task.

The initialization phase is active only at the beginning of a simulation, where external input is used to drive the *goal*[2] population of neurons to represent the vector **start**. The second phase consists of performing the task itself, and occurs once a cue is provided.

After the task has been initialized, the action selection system (consisting of the *basal ganglia* BG and *thalamus* THAL populations) switches to the process of generating word responses within the semantic system. The recurrent action selection system maintains word generation by simultaneously evaluating utilities of actions and selecting the action with the highest utility value. Table 1 shows the mapping between utility calculations and actions utilized by the action selection system. Since the external input initially sets the *goal* to **start**, the action selection system will select the first action due to its high utility value. This action will feed the vector **animal** as input to the population *cue*, and set the representation in the *goal* population to **think**. This action can be interpreted as the instruction "start listing animals".

Next, the semantic system begins to generate associations of the word *animal* within the *association network*. The connection between *cue* and the *association network* implements the transformation $\tilde{A}$, as described in the previous section.

The *association network* then represents a vector which is a linear combination of word-vectors associated with **animal**. For example, there might be a representation corresponding to the vector: 0.5***cat** + 0.4***dog** + 0.1***fish**. Coefficients represent association strengths between each individual word and the word *animal*, as derived from the association matrix A. A *winner-take-all* (WTA) mechanism within the network selects the vector with the largest coefficient, and projects it to the *response* population. In this example, the *response* popu-

---

[1]The model and data analysis source code are available at https://github.com/ctn-archive/kajic-cogsci2017.

[2]We use *italics* to refer to the name of a population of neurons or the vector that is represented by that population, which is to be inferred from the context. The **bold** font is used to refer to labels assigned to vectors representing a word. For example, *cue* · **animal** refers to the dot product between the vector represented by the population of neurons labeled "cue" and the vector corresponding to the word "animal".

lation would now represent the vector **cat**.

When a response has been generated, the action selection system selects the second action (see Table 1) due to its high utility value. This action projects the word represented in *response* (e.g., **cat**) to *cue*, simultaneously adding it to the representations stored in *response memory*. The *goal* continues to be **think**.

This process within the semantic system continues, with the action selection system selecting the second action most of the time. To prevent the same responses from re-appearing immediately, *response memory* is implemented as a neural integrator population. It projects inhibitory connections to *association network* in order to suppress representations of words previously generated as responses.

The last action with a fixed utility value of 0.4 is selected if utilities of all previous actions have evaluated to a lower value. This occurs when the system is unable to come up with a new response (e.g., the WTA mechanism takes too long to decide between two words). While rare, when this situation occurs, the action selection system sets *cue* to represent the input **animal** and the *goal* is set to **think**.

Network Parameters

Most parameters in the model have been left at their default values provided by the simulation software Nengo (Bekolay et al., 2014). Table 2 lists the major parameters in the model. Some parameter values (e.g., maximal firing rates) are selected randomly. Each time the model is run, a new set of such parameters are chosen. Such diversity in parameter settings is a first approximation of differences in cognitive processing that may occur across cortical regions of different individuals.

Results

We ran 141 simulations of the model for each of the three association matrices (Beagle, Ngram, and FAN) and compared them to human data. The number of simulations corresponds to the number of participants in the study by Hills et al. (2012). The simulations were run until the average number of responses produced matched the average number of responses given by human subjects within three minutes.

For each simulation run, we recorded word responses as decoded vector representations in the *response* population, and inter-item response times (IRT) as times between the onset of the current response and the previous response. Here we consider only relative timings (i.e., the time differences between responses), as mapping to absolute timing (i.e., exact duration of the experiment) would require consideration of the time it takes for other processes to occur, such as visual perception and motor responses, which are not part of this model.

The model responses were evaluated using the same scripts developed for the analysis of the human data, provided in Hills et al. (2012). Each response is assigned an animal category, and the clusters are identified as sequences of responses

within the same category. An animal that could be assigned to two clusters is assigned to both.[3]

The first analysis compares the pairwise similarity of a word and the words preceding it within a cluster (Figure 2A). The similarity is computed as a dot product between two BEAGLE vectors corresponding to the two words in a word pair (Hills et al., 2012). The experimental results in Figure 2A show that the word most similar to the recent word in the patch is the one preceding it, supporting the theory of locality in a memory structure. For the model, this trend is observed with the Ngram and the FAN association matrices, and less so with the BEAGLE association matrix, for which the similarity appears to have a flat trend independent of the position in the cluster.

The second analysis compares the pairwise similarity of subsequent items relative to the position of an item in the cluster (Figure 2B). Human data shows that the lowest pairwise similarity occurs at the cluster transition points, indicated by '1' on the x-axis in the figure. That point shows the similarities between the first word in a cluster and the last word in the preceding cluster. For humans, the mean similarity μ at the cluster switch is $\mu = 0.92$ with standard deviation $\sigma_\mu = 0.01$. The model using FAN data shows comparable results ($\mu = 0.93, \sigma_\mu = 0.01$). For the Ngram and the BEAGLE association matrices this effect is weakly observable ($\mu = 1.00, \sigma_\mu = 0.01$ and $\mu = 1.01, \sigma_\mu = 0.01$, respectively), as the word similarity at the transition point remains above the subject's average.

The third analysis concerns the position of a word item within a cluster and the speed of generating a word. The ratio between the average IRT for an item and the participant's mean IRT over the entire task is shown in Figure 2C. Human participants take the most time to produce the first word in a new cluster (reported $t(140) = 13.14, p < .001$) and least time to produce the second word in a new cluster (reported $t(140) = 11.92, p < .001$). This observation is the hallmark prediction of the optimal foraging strategy, suggesting that cluster switches occur when the current IRT increases over the mean IRT value. Figure 2C also shows that the model using the FAN association matrix exhibits the same effects as observed with human responses. It takes significantly more time to generate the first words in a new cluster ($t(140) = 4.78, p < .001$), compared to the second words in the cluster ($t(140) = 4.78, p < .001$).

Discussion

We have proposed a spiking neural network model that performs the semantic fluency task and shows a good match with human behavioral data. In particular, we embed association data in connections between neurons within a large recurrent network and investigate which source of association information provides the closest match to human performance. Our focus is on identifying plausible, causal neural mechanisms

---

[3]See Troyer et al. (1997) for more detailed description of the categorization procedure.

Table 2: List of model parameters

| Name	Value (unit)	Explanation
d	256	Dimensionality of word vectors
$assoc\_th$	0.3 (or 0.25)	Default WTA input threshold (Ngram, BEAGLE threshold)
c_{cs}	3	*Cue* to *association network* connection strength
c_{fs}	0.2	*Cue* feedback connection strength
c_{inh}	-5	*Response memory* to *association network* inhibitory connection strength
τ_{syn}	0.1 ms	Synaptic time constant between *association network* and *response*
τ_{syn}	0.005 ms	Synaptic time constant (default)
max_rate	200–400 Hz	Range for maximal neural firing rates (default)
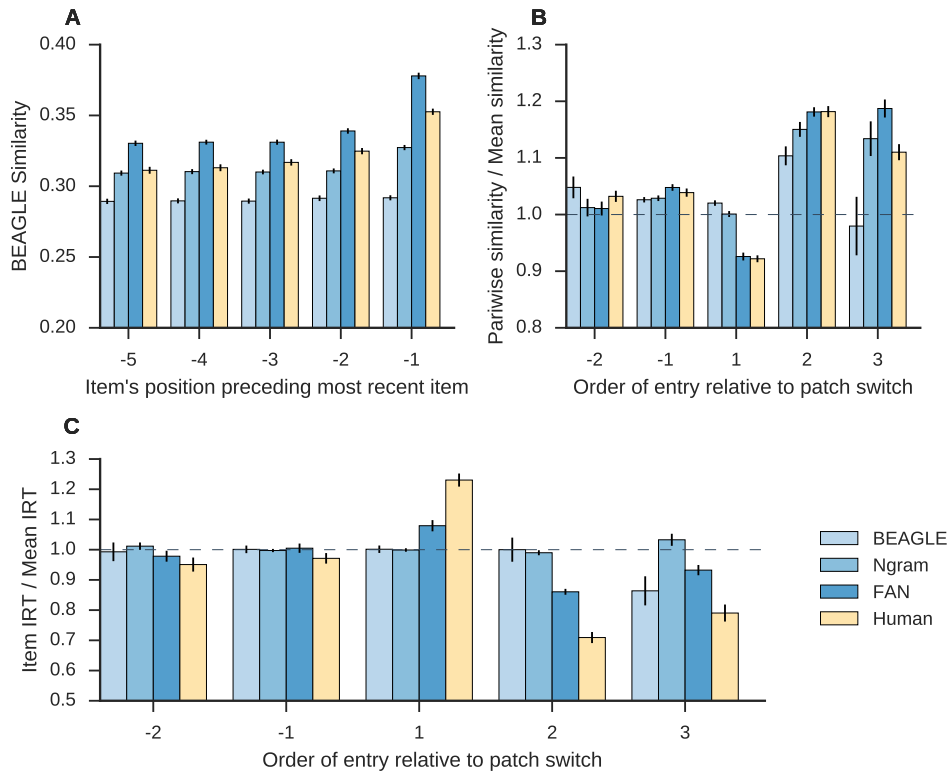
Figure 2: Comparison between model responses for FAN, Ngram and BEAGLE association matrices (blue) and human responses (yellow, reproduced from Hills et al., 2012). **A**: Pairwise similarity between a word and the words preceding it within the same categorical cluster. **B**: Pairwise similarity between subsequent words. For example, the bars above '1' indicate the relative pairwise similarities between the first item in a cluster, and the last item in the previous cluster. **C**: Inter-item response times (IRT) between subsequent words. Standard errors of the mean are shown with error bars in all plots.

for performing such tasks. To that end, we have identified computational requirements in terms of processes and relevant neural parameters, and here we discuss how they affect the model's behavior.

The model produces responses in a way that is consistent with predictions made by optimal foraging theory proposed to be used by humans (Hills et al., 2012). It is more likely to switch animal categories when the average similarity of subsequent responses drops below, or gets close to, the overall mean similarity. This effect was observed with all three association matrices, but is most pronounced with the FAN

matrix.

However, the analysis of timing effects allowed us to clearly distinguish between the three matrices. The model using FAN exhibited the same timing effects as observed with human responses. This timing effect was not observed with other association matrices (see Figure 2C). The similarity between cognitive processes involved in free association task and in the semantic fluency task (Jones et al., 2015) is a likely candidate to explain the effectiveness of free norms in matching the experimental data. However, this result could also be seen as support for the plausibility of the proposed neu-

ral mechanisms, as they are able to generate behaviors in accordance with these underlying associations. We expect that a better understanding of cognitive processes involved in free associations could aid understanding of the processes underlying semantic fluency. Our model may prove useful in exploring a variety of possible ways that such associations are neurally realized, as the direct embedding in connection weights as done here is only one possibility.

When building biologically constrained neural models, timing is a highly constrained property of a model. Here, the timing of responses is sensitive to both neural time constants and our characterization of concept representation. This is in contrast to previous models that directly use semantic networks, where timing is a separate and independent parameter. For instance, we identified that a longer synaptic time constant was needed between the *association* network and the *response* populations to stabilize the representation. This leads to the prediction that this network will be rich with NMDA receptors in the biological system. These receptors have significantly longer time constants than the more common AMPA receptors. Also, NMDA receptors can be found in the hippocampus, a brain structure located in the medial temporal lobe, whose function has been implicated in semantic and episodic memory.

Our characterization of neural concept representation also has an effect on the timing responses. Specifically, we have observed that the dimensionality of employed vector representations needed to be sufficiently large to achieve experimentally observed timing effects. While we find that $d = 256$ suffices for this purpose, a systematic search of dimensionality effects on the performance is needed to see how it affects the behavior. We have tested this model with lower values (e.g., $d = 64$) and it produced results in support of local search strategy, yet it failed to provide a good match with the timing data. In other work, we have suggested that $d \approx 500$ is necessary for representing human-scale conceptual structures (Eliasmith, 2013), which is consistent with this newer observation.

Acknowledgments

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psyc. Rev.*, *122*(3), 558.

Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T. C., Rasmussen, D., ... Eliasmith, C. (2014). Nengo: A Python tool for building large-scale functional brain models. *Front. in Neuroinformatics*, *7*(48).

Bousfield, W., & Sedgewick, C. (1944). An analysis of sequences of restricted associative responses. *The Journal of General Psychology*, *30*(2), 149–165.

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical population biology*, *9*(2), 129–136.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psyc. Rev.*, *82*(6), 407 - 428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. New York, NY: Oxford University Press.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: computation, representation, and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psyc. Rev.*, *119*(2), 431.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.

Jones, M. N., Hills, T. T., & Todd, P. M. (2015). Hidden processes in structural representations: A reply to Abbott, Austerweil, and Griffiths (2015). *Psyc. Rev.*.

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psyc. Rev.*, *114*(1), 1.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... others (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407.

Nematzadeh, A., Miscevic, F., & Stevenson, S. (2016). Simple search algorithms on semantic networks learned from language use. In *38th Cog. Sci. Proceedings* (pp. 1313–1318). Austin, TX: Cognitive Science Society.

Quiroga, R. Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nature Rev. Neurosci.*, *13*(8), 587–597.

Rissman, J., & Wagner, A. D. (2012). Distributed representations in memory: insights from functional brain imaging. *Annual Rev. of Psyc.*, *63*, 101–128.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cog. Sci.*, *29*(1), 41–78.

Thurstone, L. (1938). *Primary mental abilities* (No. 1). University of Chicago Press.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138.

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford, UK; New York: Oxford University Press.