

Experience as a Free Parameter in the Cognitive Modeling of Language

Brendan T. Johns,¹ Michael N. Jones,² & Douglas J. K. Mewhort³

btjohns@buffalo.edu, jonesmn@indiana.edu, mewhortd@queensu.ca

¹Department of Communicative Disorders and Sciences, University at Buffalo, Buffalo, NY

²Department of Psychological and Brain Sciences, Indiana University, IN

³Department of Psychology, Queen's University, Kingston, ON

Abstract

To account for natural variability in cognitive processing, it is standard practice to optimize the parameters of a model to account for behavioral data. However, variability reflecting the information to which one has been exposed is usually ignored. Nevertheless, most language theories assign a large role to an individual's experience with language. We present a new way to fit language-based behavioral data that combines simple learning and processing mechanisms using optimization of language materials. We demonstrate that benchmark fits on multiple linguistic tasks can be achieved using this method and will argue that one must account not only for the internal parameters of a model but also the external experience that people receive when theorizing about human behavior.

Keywords: Cognitive modeling; Model optimization; Language processing; Corpus-based models.

Introduction

Models of cognition often have to deal with troublesome sources of variance that other fields (e.g., physical systems) do not. For example, no two individuals process a stimulus in the same way, and the same individual rarely processes the same stimulus identically at multiple times. In addition to individual differences and temporal stability, there is true random and measurement variance. While many of the sources of variance can be represented by free parameters, much of what may be systematic variance ends up being encapsulated by an overall noise parameter, often thought to reflect the inherent stochastic nature of the response process (Shiffrin, Lee, Kim, & Wagenmakers, 2008).

Almost every cognitive model contains free parameters, coefficients that are initially unknown, but are estimated from the observable data. The exact values for free parameters do not change the model's architecture—the theory that the model formalizes should be independent of its parameter values—but the settings do change a model's behavior. Hence, researchers use estimation methods to find the set of parameters that maximize a model's fit to data, and those parameter estimates are often allowed to vary across different data sets to which the model is applied.

A tacit assumption in cognitive models is that behavioral differences across individuals or tasks can be explained by differences in process parameters. But an alternative source of variance, often ignored, comes from differences in the subject's individual learning history or variance in memory representations selected for a task, independent of changes in the process parameters.

Theories of cognition commonly assume that aspects of the external world are stored internally. The storage assumption

applies to memory (Anderson & Schooler, 1991), perception (Barsalou, 1999), and linguistic organization (Landauer & Dumais, 1997). In effect, the assumption acknowledges that human beings are embedded in a structured physical environmental that informs learning and that constrains behavior. As Simon (1969; p. 53) makes clear, "The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves".

Although we acknowledge natural variation in processing mechanisms, we explore in this paper the other source of variation—the environmental information to which people have been exposed—on lexical tasks. Subjects differ in what they know, and the differences should cause a corresponding change in behavior. Of course, the effect of variable knowledge depends on the specific task. For example, accumulated linguistic knowledge likely affects a lexical decision experiment more than a non-verbal perceptual identification task, so including linguistic knowledge when modeling lexical decision makes a good deal of sense.

One way to build linguistic information into a model is to use a representation of word meaning constructed from a standard corpus, such as the TASA corpus (introduced by Landauer & Dumais, 1997). The TASA corpus includes paragraphs from textbooks, from grades 1 to 12 and has been used as the gold standard in tests of co-occurrence models (e.g., Landauer & Dumais, 1997; Jones & Mewhort, 2007); it has frequently been integrated into processing models in cognate areas (e.g., Johns, Jones, & Mewhort, 2012). Although the TASA corpus is likely representative of the linguistic experiences that subjects have experienced, it is not intended to map exactly onto the experiences of specific individuals. Indeed different groups of subjects may have had exposure to wildly different linguistic sources, depending on culture, geography, educational system, and so forth. Hence, for any group of subjects, there is a natural variation in their knowledge, variation that should impact the behavior of those subjects on specific laboratory tasks.

Paradoxically, most theorists recognize that knowledge is central to performance in standard laboratory tasks but rely on a single corpus to model cognition. By relying on a single corpus, they ignore an important source of variability, namely the different knowledge that individuals bring to the task. If one could estimate a group of subjects' average linguistic experience, it should be possible to account for behavioral data at a more refined level.

The present article describes a new method for taking into account both the internal parameters of a model as well as the environmental information that defines a subject's unique

experience. Unfortunately, one cannot track a subject's linguistic history. As a proxy of that history, however, we collected a very large diverse set of language materials and combined it with a simple mechanism to uncover the most informative texts for a specific set of data.

To demonstrate the generality of our method, it was applied to both lexical organization and lexical semantic data. Modelling each of these tasks requires information to be accumulated into the mental lexicon, after which it will be acted upon to accomplish a specific task. The aim is to show that, by combining experiential fitting with realistic cognitive models across multiple areas, benchmark accounts of language-based behaviors can be attained.

Corpora and Data Fitting Methodology

To estimate the type of linguistic information used in a certain behavioral task, we use a wide variety of large language sources. These sources are then split into smaller sections, and it is iteratively determined which sections maximized the fit of a model to a set of data. At the end of the iterative process, the algorithm will have determined the most informative set of texts needed to explain performance on the task. Here, we describe the training materials and specifics of the method used to fit the models.

Training Materials

The texts come from five different sources: (a) Wikipedia (Shaoul & Westbury, 2010), (b) Amazon product descriptions (attained from McAuley & Leskovec, 2013), (c) 1,000 fiction books, (d) 1,050 non-fiction books, and (e) 1,500 young-adult books. All of the books were attained from e-books, and the vast majority were written in the last 50 years by popular authors. The set of sources—from an online encyclopedia to books targeted at young adults, to marketing materials for a large range of products—was designed to represent a broad set of possible experiences that an individual might have with written language. It is impossible, of course, to span the entire range of possible linguistic information, but the present materials represent a substantial range of texts, one that should give experiential fitting a fair test. To equate each source's contribution, each was trimmed to six million sentences, for a total of 30 million sentences across all texts (approximately 400 million words).

The data-fitting method will determine which set of texts is the most informative for fitting a particular experiment, just as statistical methods are used to estimate the optimal free parameters of a model. The corpora were split into small sections of 50,000 sentences yielding 120 sections for each corpus, for a total of 600 different sections across the corpora. Each section is large enough to allow for a measure of how much linguistic information the section contains, but is small enough that the different sections can still be combined to determine an optimal set of language.

Data Fitting Methodology

The goal of the data-fitting algorithm is to determine the combination of the sources that gives the best fit for a specific

model to a set of data. To do so, we used a hill-climbing algorithm iteratively to select the sections that maximize the model's likelihood of generating various behavioral datasets.

A hill-climbing algorithm is an iterative local search algorithm, where a model is fit by incrementally improving its fit to a set of data. Once an increase in fit is no longer possible, the algorithm terminates. For experiential fitting, the first iteration selects the section that provides the best fit. Subsequent iterations add additional sets on top of the previously selected sections, to construct an overall training corpus. Once a section has been selected, it can no longer be used (sampling without replacement). Hence, the training materials increase their resolution continuously, in correspondence with the structure of the set of data attempting to be modeled. Fitting ends when the addition of a further section into the overall corpus does not increase the fit of the model to the data. To avoid getting stuck on local maxima, 10 unique starting points were made, in a rank order of the best fitting sections. The best fit will be displayed in the below simulations.

Discussion

To explore the power of experiential fitting, a large amount of text was assembled across a number of different sources. To determine the optimal set of linguistic data to explain a set of data, the texts were split into smaller pieces, and a hill-climbing algorithm was used iteratively to find the selection of text that maximally increased the fit of a model to a set of data. One could think of the process as a kind of parameter fitting (see Shiffrin, et al., 2008), but instead of optimizing the internal parameters to explain a set of behavioral data, the procedure optimized the structure of the external world (i.e. linguistic information). Optimizing the linguistic information allows us to determine the power gained by accounting for the variance in linguistic experience to an explanation of human behavior. That is, if linguistic behavior is related to the structure of linguistic experience, determining the optimal set of language materials with which to train a model should provide a substantial increase in the fit of the model.

Lexical Semantics

Models of semantic memory, particularly Latent Semantic Analysis (Landauer & Dumais, 1997), have strongly influenced studies of the effect of linguistic experience. LSA showed that a simple averaging mechanism, when combined with sufficient amounts of language information (derived from a large text corpus), can construct a representation of the meaning of words that is closely matches how people use language.

The model used here is derived from BEAGLE (Jones & Mewhort, 2007), a random vector accumulation model. The BEAGLE model is based on using sentential information in the learning process. In this model, words are represented by two vector types: a static environmental vector, that represents the perceptual (visual/auditory) aspects of a word, and dynamic context/order vectors, which mark both co-

occurrence and simple syntactic usages of a word. Each time a word is seen in a corpus of text, the dynamic vectors are updated. For context information, updating is done by summing the environmental vectors of the other words that occurred in the sentence with it (with high frequency function words removed). Accordingly, the context representation accumulates pure co-occurrence information. Order vectors, by contrast, accumulate rudimentary syntactic information, by recording the position of words that surround the usage of a word. The original BEAGLE model used circular convolution to form an n-gram representations of a sentence. Here, we will use a simplified form of the model (see Recchia, Sahlgren, Kanerva, & Jones, 2015), because the simplified form is less computationally expensive. We refer the reader to Recchia, et al. (2015) for a complete description of the simplified form of BEAGLE. In the following simulations, order, context, and the complete (the sum of the order and context vectors) representations were used.

The model was tested using two different data types: (a) synonym tests, and (b) item-level semantic priming. Following Landauer and Dumais (1997), synonym tests have become a standard test for models of semantic representation. In the synonym test, subjects are required to pick the word from a set of four that is most similar in meaning to a target word. A real-world example is the Test of English as a Foreign Language (TOEFL). Landauer & Dumais used 80 questions from the TOEFL and reported that LSA achieved an accuracy of 55% on this test.

Semantic priming will also be analyzed, a type of data that semantic space models have had success in accounting for (Jones, Kintsch, & Mewhort, 2006). In behavioral experiments, subjects are asked to perform simple tasks, such as lexical decision, but the target word is preceded by a prime. The prime can be semantically related or not, and the benefit provided by the prime is measured in terms of the speedup seen when the target is preceded by a semantically related item versus a semantically unrelated word.

Hutchison, Balota, Cortese, & Watson (2008) have shown that models of semantic representation may succeed at the mean level across items but fail at the individual word level. They examined priming in lexical decision for 300 different items and found that semantic variables offered minimal fits to the data, with forward association strength having the best correlation to overall levels of priming at $r = 0.164$, $p < 0.01$, while LSA had a non-significant correlation of $r = 0.053$. Clearly semantic priming data are difficult to account for at an item level; hence, the data provide an excellent test for the power of experiential fitting.

Data Fitting Methodology

As noted earlier, all corpora were split into sections of 50,000 sentences and vector sets were generated for all the three types of information created by BEAGLE (context, order, and complete). The hill-climbing algorithm selected semantic vectors to maximize the model's performance on the TOEFL test, and rank correlation in semantic priming. That is, the necessary semantic information was refined iteratively to maximize the fit to the data. Iterative refinement halted when

adding new material failed to increase the quality of the fit. To minimize the chance of falling into a local minimum during the fit, 10 random starts were used. To form a comparison, 50 resamples of the full corpus (of 30 million sentences), and the average performance increase across each 50,000 section of this corpus was recorded. This will provide a measure of how successful the model is independent of experiential optimization.

Results

The top panel of Figure 1 shows accuracy on the TOEFL test as a function of the number of sentences included in the fit; it shows results for the three kinds of information (item, order, and complete) along with a control condition in which the sections of text were assembled randomly. For the random corpora, the results were concatenated at 10 million sentences in order to aid in visualization. The complete model achieved the best performance at 97% accurate at only 1.1 million sentences. The context representation maximized at 92% accurate at 3 million sentences, while the order representation maximized at 82% accurate at 2.1 million sentences. For the random corpora, the average maximum performance was 57%, consistent with the past results (Jones & Mewhort, 2007). The complete representation is essentially performing at the same level as a native English speaker, an impressive level of performance for a rather simple model.

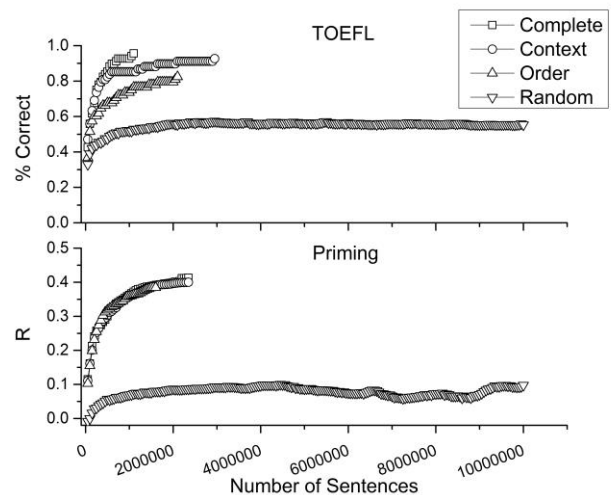


Figure 1. Results of BEAGLE and experiential optimization.

We also examined Hutchison, et al.'s (2008) item-level semantic priming results. Recall that it has been challenging for semantic-space models to account for item-level results. Figure 3 shows the fitted correlation as a function of the number of sentences for item, order, complete (combined) and random controls. As is shown in the bottom panel of Figure 1, all three representation types (item, order, and combined) provided a good fit to the item-level data in semantic priming. There was not a great deal of difference among the non-random representations, the complete model did offer the best fit at $r = 0.412$, $p < 0.001$. Note that the complete model provided a better fit than all the semantic

variables tested by Hutchison, et al. (2008). Indeed, it approached the fit of their 18-variable regression ($r = 0.5$).

Discussion

Semantic space models have been fundamental in exploring the influence of the linguistic environment on human behavior. This section explored the power that comes from combining a simplified form of a popular semantic space model (Jones & Mewhort, 2007) with experiential fitting, with the result being benchmark fits for every dataset analyzed. What this suggests is that the representations that people form in semantic memory are heavily influenced by the content of experience, and by constructing corpora that reflects this experience, better representations can be constructed.

However, one question about this method that needs to be determined is what source of variance is exploited in these simulations. It needs to be shown that the method is sensitive to group characteristics, where groups of subjects who have likely had different linguistic experiences, are found to have different corpora statistics by the experiential fitting method. That is, the method does not just exploit noise in the different datasets, but is actually approximates the type of experiences a group of subjects may have had.

Lexical Organization

A prominent area in the study of word recognition has focused on examining the influence of environmental variables on the retrieval of words from the mental lexicon. Classically, word frequency has been the most important lexical variable used to examine lexical retrieval, based on findings that higher frequency words are processed more efficiently. This has led to word frequency to be considered a central information type to models of lexical retrieval.

The exact nature of frequency effects has recently been questioned on several grounds. In one line of research, Adelman, Brown, and Quesada (2006) demonstrated that a measure that builds a word's strength in memory by counting the number of contexts that a word occurs in (operationalized as the number of document occurrences across a corpus) provides a superior fit to retrieval times than word frequency; this finding has been replicated across different corpora and datasets (Adelman, et al., 2006; Brysbaert & New, 2009). This measure is commonly referred to as a word's contextual diversity (CD).

However, Adelman et al.'s (2006) document count measure ignores the semantic diversity of the contexts that a word occurs in. To examine this possibility more closely, Jones, Johns, and Recchia (2012) conducted an artificial language learning experiment that manipulated word frequency and contextual diversity, such that certain words occurred with different sets of words (high semantic diversity), while others repeatedly occurred with the same set (low semantic diversity). Although there was no effect of diversity for low-frequency words, high frequency words were retrieved more quickly when they had been learned across multiple diverse contexts, indicating that processing savings occurred only

with a change in context. On the basis of these results, and a corpus analysis, Jones, et al. (2012) proposed a new model that builds a more accurate measure of a word's strength in memory, entitled the semantic distinctiveness memory (SDM) model.

The SDM builds a word's strength in memory by weighting each new context by how much unique information that context provides about the meaning of the word. Across various corpora, this model was able to account for a larger amount of variance to a mega dataset of lexical decision and naming times over word frequency and a document count. Additionally, Johns, et al. (2012) demonstrated that the advantage for a semantic diversity count extends to spoken word recognition performance. Johns, Dye, and Jones (2016) have extended the results of the artificial language experiment of Jones, et al. (2012) with natural language materials and found similar results.

The simulations in this section will compare WF, CD, and SDM magnitudes, in combination with experiential fitting. The main source of data is 40,000 lexical decision times attained from the English Lexical Project (ELP; Balota, Yap, Cortese, Hutchison, Kessler, Loftis, Neely, Nelson, Simpson, & Treiman, 2007). This is a standard dataset that has been used to differentiate different lexical information sources (Brysbaert & New, 2009; Jones, et al., 2012). Additionally, a set of 2,900 lexical decisions times for young and old adults attained from Balota, Cortese, and Pilotti (1999) was used to test the sensitivity of the experiential fitting method to different subject groups.

Data Fitting Methodology

Because the SDM uses paragraphs, the fitting method split each corpora into 3,000 paragraphs/documents (roughly equivalent to 50,000 sentences). For the Wikipedia corpus, this was a single document in the encyclopedia. For the Amazon product descriptions, one product description was considered a separate document. For the books, due to how they are formatted, there was no simple method to split them into paragraphs. Instead a moving window, with a size of 15 sentences, was used to assemble paragraph-like units.

Typically, the SDM is trained on a whole corpus, as the model is dynamic: previously experienced information is used to determine what should be stored for any new context. However, the model is quite computationally complex, so magnitudes were derived separately for each section. Overall magnitudes were then the sum of the different selected sections, which was also done for the WF and CD variables. These variables were transformed with a natural logarithm before assessing the correlation to the data.

Results

The results of the experiential fitting method on the z-transformed ELP lexical decision time data are displayed in the top panel of Figure 2. Only the results of the SDM are displayed in this figure, because all three measures produce similar results (explored further below). This result is contrasted with the fit that CD values from the SUBTLEX corpus (Brysbaert & New, 2009) provides for this data set, as

it provides the best fits to this data currently available. The figure demonstrates that the use of experiential fitting allows for a large increase in fit for retrieval latencies, even when compared against a very well-constructed corpus, as it outperformed SUBTLEX by a large margin. Additionally, the randomized corpora also achieved a correlation that equaled the SUBTLEX corpus (Brysbaert & New, 2009), demonstrating that the source materials that the experiential fitting method was using was of very high quality.

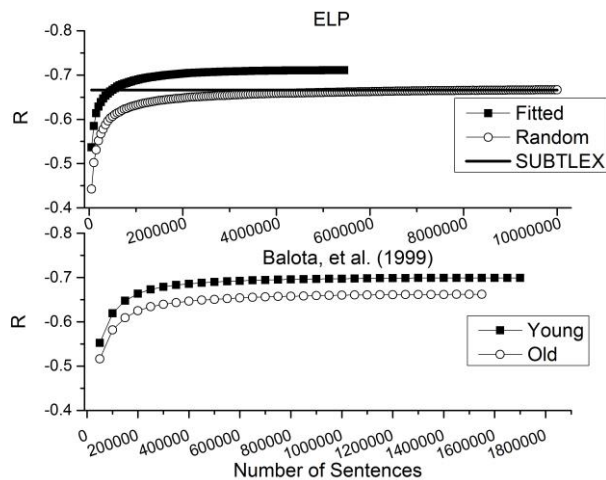


Figure 2. Results of SDM with experiential optimization.

As has been pretty previously shown, magnitudes from the SD model had the highest correlation, with an $r = -0.708$, $p < 0.001$, compared with an $r = -0.702$, $p < 0.001$ for contextual diversity, and $r = -0.701$, $p < 0.001$ for word frequency. As a comparison, the correlation for CD values from SUBTLEX is an $r = -0.666$, $p < 0.001$. The SDM model providing the superior fit is consistent with past results (Jones, et al., 2012; Johns, et al., 2012), but the interesting aspect of this simulation is the power that experiential fitting provided to all three variables.

As noted previously, there is still a question of what the source of variance that the method is capitalizing on, as it is possible that it is not capitalizing on group or individual characteristics, but instead random noise within the different datasets. As a test of this, 2,900 lexical decision times were attained from Balota, et al. (1999) for younger and older adults. The bottom panel of Figure 2 displays the fits to Balota et al.'s data for the SD model, and demonstrates that a high level of fit was attained for both subject groups, but with a higher fit to younger than older subjects, a standard finding. However, a more interesting analysis is to examine the composition of the resulting corpora for the two subject groups. To do this, the proportion of the different sources that was selected was recorded across 20 runs of the hill-climbing algorithm. These runs were done by removing the previously selected first section for the current run, so that each run begins differently. The results of this analysis are contained in Figure 3.

There was no difference in proportions selected for the non-fiction, Wikipedia, and Amazon sections, but there was a

highly significant difference for the young adult sections [$F(1,39) = 203.51$, $p < 0.001$] and the fiction sections [$F(1,39) = 219.45$, $p < 0.001$]. These differences emerge because the young subject group had a higher proportion of young adult sections, while older adults were better described by the fiction sections. Given the composition of the different corpora, this suggests that the retrieval time data of these different groups are sensitive to the statistics of different linguistic sources that the subjects have experienced: young adults are better described by simpler examples of language as encoded in young adult books, but older adults are better accounted for by more linguistically diverse fiction and literature books. At least anecdotally, this is consistent with the type of linguistic experiences these subjects likely had.

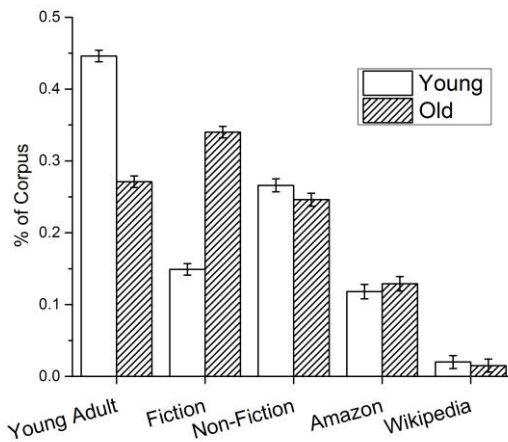


Figure 3. Proportion of different sections selected for young and old subjects.

Discussion

This section demonstrates that the use of experiential fitting can be expanded easily to examine lexical retrieval. There is a rich history of using environment variables (i.e. word frequency) to examine word retrieval patterns, with recent research pointing to the importance of contextual and semantic variables in the construction of a word's strength in the mental lexicon (Adelman, et al., 2006; Jones, et al., 2012). It was found that the SDM model, previously shown to provide a superior fit to large scale lexical decision data than word frequency or a document count, when combined with experiential fitting, provides a better accounting than previously published norms. Additionally, in an examination of young and older adult lexical decision data (Balota, et al., 1999), it was found that the method was sensitive to group characteristics, suggesting that the method is fitting to the experiences that a group of subjects may have had with language.

General Discussion

The current article describes a new method for optimizing cognitive models through experiential fitting, where the

information that a model “knows” is manipulated to provide the best fit to a set of data. The manipulation was done by assembling very large sets of texts spanning multiple areas, including an online encyclopedia, product descriptions from Amazon, and sets of fiction, non-fiction, and young adult books. These corpora were split into small sections, and a hill-climbing algorithm was used to determine the best combination of these materials for a specific model and set of data. It was demonstrated that this method, combined with experience-based cognitive models, provided benchmark fits to multiple types of lexical information.

The underlying philosophy of our method is similar to standard parameter fitting methods (Shiffrin, et al., 2008), which assume that there is natural variability in the parameters that define the cognitive processes that underlie behavior. Similarly, experiential fitting is designed around the idea that there is natural variability in the knowledge bases that different subjects groups have (and also in individual subjects) that leads to variability in behavior.

One of the exciting aspects of this technique is that it provides a mechanism by which to discriminate the varying contributions of internal cognitive mechanisms and external information, an old goal in the cognitive sciences (Anderson & Schooler, 1991; Simon, 1969). If one accepts that language is dictated by a complex interaction of biological and cultural evolution (Christiansen & Chater, 2008), then it is necessary to determine how much of the complexity in human behavior is derived from evolved mechanisms in the brain and how much is provided by the heavily structured environment in which humans are embedded. The simulations reported here provide substantial evidence that the information used to train a model is very important to a model’s behavior, just as human behavior is sensitive to the knowledge that a person has gained. The simulation reported in Figure 3, where the experiential fitting method found different corpus constructions to explain younger and older adult’s lexical decision data is a promising first step that group-level experiences can be estimated with this method.

More generally, this work points to the usefulness of building cognitive models around a learning mechanism that is capable of extracting information from large text-bases, an issue that has been explored in greater detail elsewhere (e.g. Johns, Jones, & Mewhort, 2012). By basing a model’s performance in the learning of large-scale environmental information, it provides a stronger case for the plausibility of a model, as it is capable of scaling to levels of data input that a typical person may receive.

As Simon (1969) described, in order to provide a complete account of behavior, it is necessary to understand both the internal mechanisms and the environmental information that people use to behave. This is especially important in the study of language, as the vast majority of psycholinguistic theories have focused on the internal mechanisms that are responsible for linguistic behavior, while the influence of environmental information has been downplayed.

Downplaying environment information was necessary in early work because we lacked both large amount of texts and

computational resources, but neither of these factors are limitations anymore. It is readily possible to examine the impact of linguistic information on human behavior, and by optimizing the linguistic information to which a model is exposed, it provides a powerful test of a model’s ability to account for behavioral data.

References

- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision time. *Psychological Science*, 17, 814-823.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396-408.
- Balota, D.A., Cortese, M.J., & Pilotti, M. (1999). Item-level analyses of lexical decision performance: Results from a mega-study. In *Abstracts of the 40th Annual Meeting of the Psychonomics Society* (p. 44). Los Angeles, CA: Psychonomic Society.
- Balota, D. A., Yap, M.J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, 39, 445-459.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977-990.
- Christiansen, M., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31, 489-558.
- Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2008). Predicting semantic priming at the item level. *The Quarterly Journal of Experimental Psychology*, 61, 1036-1066.
- Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2012). A synchronization account of false recognition. *Cognitive Psychology*, 65, 486-518.
- Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *Journal of the Acoustical Society of America*, 132, EL74-EL80.
- Johns, B. T., Dye, M., Jones, M. N. (2016). The influence of contextual variability on word learning. *Psychonomic Bulletin and Review*.
- Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, 55, 534-552.
- 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., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, 66, 115-124.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato’s problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- McAuley, J., and Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM Conference on Recommender Systems (RecSys)*, 165-172.
- Recchia, G. L., Sahlgren, M., Kanerva, P., & Jones, M. N. (2015). Encoding sequential information in vector space models of semantics: Comparing holographic reduced representation and random permutation. *Computational Intelligence & Neuroscience*. doi.org/10.1155/2015/986574
- Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEX. *Behavior Research Methods*, 42, 393-413.
- Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E. J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, 32, 1248-1284.
- Simon, H. A. (1969). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.