

# The influence of contextual diversity on word learning

Brendan T. Johns<sup>1</sup> · Melody Dye<sup>2</sup> · Michael N. Jones<sup>2</sup>

© Psychonomic Society, Inc. 2015

**Abstract** In a series of analyses over mega datasets, Jones, Johns, and Recchia (Canadian Journal of Experimental Psychology, 66(2), 115–124, 2012) and Johns et al. (Journal of the Acoustical Society of America, 132:2, EL74-EL80, 2012) found that a measure of contextual diversity that takes into account the semantic variability of a word's contexts provided a better fit to both visual and spoken word recognition data than traditional measures, such as word frequency or raw context counts. This measure was empirically validated with an artificial language experiment (Jones et al.). The present study extends the empirical results with a unique natural language learning paradigm, which allows for an examination of the semantic representations that are acquired as semantic diversity is varied. Subjects were incidentally exposed to novel words as they rated short selections from articles, books, and newspapers. When novel words were encountered across distinct discourse contexts, subjects were both faster and more accurate at recognizing them than when they were seen in redundant contexts. However, learning across redundant contexts promoted the development of more stable semantic representations. These findings are predicted by a distributional learning model trained on the same materials as our subjects.

**Keywords** Language comprehension · Semantic memory · Word recognition · Computational modeling

---

✉ Brendan T. Johns  
btjohns@buffalo.edu

<sup>1</sup> Department of Communicative Disorders and Sciences, University at Buffalo, 122 Carey Hall, Buffalo, NY 14214, USA

<sup>2</sup> Department of Psychology, Indiana University, Bloomington, IN, USA

## Introduction

The notion that word frequency is a principal variable in how words are processed has been recognized in the psychological literature for more than half a century (Broadbent, 1967; Howes & Solomon, 1951). Frequency has proved to be a robust predictor of performance across a wide variety of tasks. For instance, high frequency words show a uniform advantage in perceptual and production tasks, with shorter response latencies and higher accuracy in tests of perceptual identification (Morton, 1969), word naming (Forster & Chambers, 1973), and lexical decision (Scarborough, Cortese, & Scarborough, 1977).

Accordingly, many models of lexical access are built on the assumption that repetition is key to entrenchment in memory, such that the more times an item is encountered, the more easily it will be processed or accessed. This *principle of repetition* is often formalized as a mental counter, which may “bias” detection of an item (lowering its resting state threshold – Morton, 1969; or raising its baseline activation level – Coltheart et al., 2001), or which may increase its accessibility in a serial access system (Murray & Forster, 2004).

However, important findings have called into question the extent to which pure repetition matters, independent of other factors. A key confound is environmental: High frequency words will not only have been experienced more often, but are also likely to have been experienced more recently (Scarborough et al., 1977), and in a greater variety of contexts (Dennis & Humphreys, 2001). Words that are spread more evenly across contexts exhibit distinct properties from those that cluster more densely (Church & Gale, 1995), and these differences appear to have important consequences for processing.

A word's *contextual diversity* – that is, the number of different contexts in which it appears – significantly influences how that word is learned and remembered. Words that are

present in a greater diversity of contexts are acquired more rapidly in early learning (Hills et al. 2010) and are processed more quickly and accurately in naming and lexical decision (Adelman, Brown, & Quesada, 2006; Schwanenflugel & Shoben, 1983). Likewise, in standard episodic memory tasks high diversity benefits recall (Lohnas, Polyn, & Kahana, 2011) but impairs recognition (Steyvers & Malmberg, 2003). The influence of contextual diversity has also been linked to the benefit of spaced over mass practice (Verkoeijen, Rikers, & Schmidt, 2004).

These empirical findings align well with the theoretical proposal that the contents of memory are organized in such a way that needed information can be accessed quickly and reliably. According to the *principle of likely need* (Anderson & Milson, 1989), the accessibility of an item in memory is not simply a function of its current match to a retrieval probe, but is also strongly influenced by its history of use. Items that have previously been retrieved in a variety of different contexts are more likely to be needed in the processing of a yet-unknown future context; hence, they should be easier to access.

However, it is still an open question how best to characterize a word's contextual diversity. The most common operationalization of the variable is to count the number of distinct documents in which a word occurs across a text corpus (e.g., Adelman et al., 2006). Recently, Jones, Johns, and Recchia (2012) demonstrated how a more nuanced measure of contextual diversity, which they termed a *semantic distinctiveness* count, provided a better fit to human word recognition latencies above and beyond pure frequency or document count (see also Hoffman, Ralph, & Rogers, 2013). This continuous measure scores a word that has appeared in multiple semantically distinct contexts more highly than one that has occurred in more redundant contexts, even when the two are balanced on both document and frequency counts. In short, a word's occurrence is weighted relative to the information overlap between the current context and the previous contexts in which it has occurred. This makes the measure dynamic: the value for a specific document depends on how much new information it contributes about the word beyond what has previously been encountered.

Across various corpora and datasets, the semantic distinctiveness count has been shown to provide better fits to visual lexical decision and naming data (Jones et al., 2012) and spoken word-recognition accuracy (Johns et al., 2012). The variable also explains a key interaction in an artificial language experiment (Jones et al., 2012), which cannot be explained by raw frequency: Repeated presentations of a word at learning only benefits subsequent processing speed if the presentation is accompanied by a change in context, a pattern also observed in Balota et al.'s (2007) mega-database. Results such as these demonstrate the importance of event history in learning, indicating that redundant experiences are not encoded as strongly as unique experiences.

That said, frequency still plays into these effects. In an analysis of words from the English lexicon project, Jones et al. (2012) found that for low frequency words, there is little effect of diversity. However, high frequency words were shown to be processed more efficiently when a word occurred in more semantically variable contexts. The reason for this is unlikely to be mere repetition. Rather, frequency is a necessary condition for variability to exist. Compared to their high frequency counterparts, lower frequency words have a more limited event history, and hence are less likely to have been sampled as broadly.

In light of these findings, Johns, Dye, and Jones (2014) proposed a model of lexical processing that captures the effects of semantic distinctiveness within a classic distributional model of lexical semantics. Distributional models (e.g., Landauer & Dumais', 1997, LSA) have been very successful at explaining semantic similarity among words as a function of their co-occurrence across documents in large text corpora. While the mechanisms of the various models have considerable theoretical differences (see Jones, Willits, & Dennis, 2015, for a review), they all construct vector representations for words based on frequency of occurrence across documents. Two words are semantically similar to the extent that they have similar covariation patterns across documents. Hence, semantically similar words like *dog* and *cat* will develop more similar vector patterns than will unrelated words.

But similarity only considers a word vector's *phase*; *magnitude* is also an important property of these vectors. The magnitude is produced by summing the elements of the vector; if the vector is simply occurrence frequency across documents, then the magnitude will equal word frequency. Hence, lexical availability (magnitude) of single words and semantic similarity (phase) between words are intricately tied together in distributional models, which can thus potentially explain both behavioral variables.

Johns et al.'s (2014; also, Jones et al., 2012) Semantic Distinctiveness Model (SDM) is a distributional model which incorporates an attention-weighting mechanism when encoding a new context entry in a word's vector. In particular, the model compares a new context that a word occurs in to a prediction of its meaning from the memory vector that has encoded its previous contexts. If the new context is congruent with the expected meaning in memory, it is encoded at a weaker intensity than if the new context is surprising.

Across various corpora, the SDM is able to account for a larger amount of variance from a mega dataset of lexical decision and naming times as compared to word frequency or a raw context count – an advantage that extends to spoken word recognition (Johns et al., 2012). In addition, Johns and Jones (2008) found preliminary evidence that encoding contexts in this fashion also provides a better fit to semantic similarity ratings. In short, SDM appears to produce vectors with both phase and magnitude that better explain human behavior across lexical access *and* semantic similarity tasks.

The broad array of results attesting to the importance of semantics on lexical access suggests that word retrieval and word meaning are based on the same environmental information, and that there is a high degree of interaction between the two systems. The preliminary data from the SDM suggest that it has the potential to explain both kinds of behavioral data and may offer a mechanistic understanding of how they are related to the statistical structure of the language environment.

To test the validity of this assumption, and to further extend the results of Jones et al.'s (2012) artificial language experiment, a novel experimental paradigm was developed to assess whether the SDM accurately captures how discourse variability at encoding influences subsequent lexical access and semantic similarity.<sup>1</sup> In training, subjects read and rated short passages containing pseudowords. Some words were encountered in highly distinctive contexts, while others were encountered across very similar contexts. Following incidental exposure at reading, subjects completed a pseudolexical decision task<sup>2</sup> (PLDT) and a semantic similarity judgment task. When trained on the same material as our subjects, the SDM predicted that whereas diverse contexts should strengthen memory for novel words, leading to faster and more accurate recognition judgments, uniform contexts should support the development of more stable semantic representations.

## Method

### Participants

Ninety-one undergraduate students at Indiana University participated in the experiment for US\$10. All were native American English speakers. Data from four subjects were discarded: two because they did not complete the experiment, and two because their performance fell below chance on the PLDT.

### Materials

The study was designed to assess how representations of novel words develop over reading multiple passages. Accordingly, ten target words were selected, all of which were low frequency and attested in a variety of discourse contexts.

<sup>1</sup> Although the models in Jones et al. (2012) and Johns et al. (2014) are theoretically equivalent, the Johns et al. sparse vector version is much more computationally efficient and can be scaled up to very large word corpora. Hence, the Johns et al. version is used to make all predictions here.

<sup>2</sup> This is the standard task label that has been used in this literature (e.g., Nelson & Shiffrin, 2013; Jones, et al., 2012). However, it is possible to conceptualize this task as an episodic memory task, as the words may not have received enough repetitions to enter the mental lexicon. A more expansive discussion of this issue is contained in the General Discussion.

Training materials were drawn from natural real-world contexts in which these targets occurred. For each target, two distinct sets of passages were developed: one set comprising five passages from a single discourse topic (low variability) and the other comprising five passages spanning a number of distinct topics (high variability). Passages were excerpted from reputable fiction and non-fiction sources, and selected such that length and semantic overlap were kept constant across targets within each condition. In addition, passages were manipulated to be similarly informative about target meaning.

However, using real word forms in training would make it difficult to separate learning at study from prior learning. To minimize the effects of pre-experimental exposure, each target was randomly replaced with a pronounceable pseudoword at the beginning of the experimental session. These replacements were drawn from a list of 20 pseudowords, which had been selected from the English Lexicon Project (Balota et al., 2007), and matched on number of letters, orthographic neighborhood size, bigram count, and reaction time and accuracy in PLDT.

### Procedure

Participants were told that they were reading standardized testing materials for clarity and comprehensibility. During the study phase of the experiment, each passage was displayed on screen for a minimum of 10 s, after which a rating scale appeared. Subjects were instructed to make a rating on a scale of 1 to 7 assessing how well they understood the passage, with 1 indicating that they did not understand it at all, and 7 indicating that they understood it perfectly. No time limit was imposed. After the subject's rating had been submitted, the passage and scale disappeared, and the program advanced to the next trial. Figure 1 depicts a sample study trial.

The study was designed such that each target word had both a uniform (low variability) and a diverse (high variability) set of passages associated with it, each of which comprised five short paragraphs. At the beginning of study, the program randomly assigned half of the targets to the uniform condition, and half to the diverse condition. Each target was then randomly assigned a pseudoword, which replaced the target across all the passages in which it occurred. Subjects

Schizophrenia is a common disorder with a prevalence of approximately 1 per cent. The illness often develops in young adults, who were previously normal, and includes hallucinations and delusions, and symptoms such as severely inappropriate emotional responses, a disorder of thinking and concentration, erratic behavior as well as social and occupational deterioration. This suggests that the disease is characterized by a covella of indicators.

**Fig. 1** A screen capture of a sample trial during the study phase. The pseudoword in this paragraph is *covella*, replacing the target word *constellation*

read a total of 50 paragraphs (ten targets  $\times$  five paragraphs), with the order of presentation randomized.

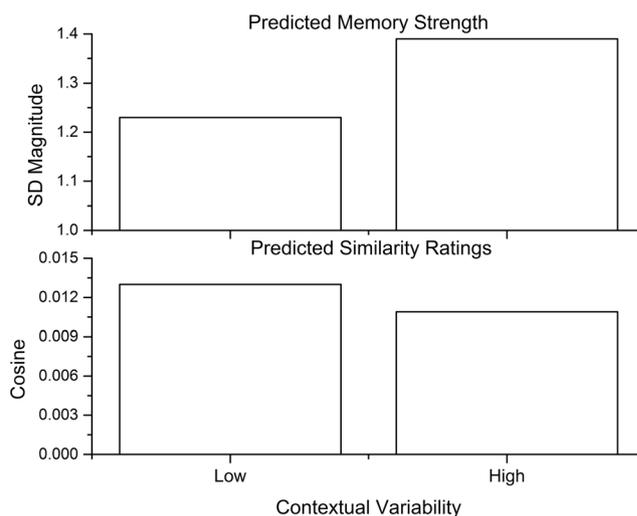
After training, subjects completed a surprise PLDT. For each pseudoword presented at test, subjects were asked to determine whether they had seen that word at study, responding as quickly and accurately as possible. Each pseudoword was preceded by a fixation cross that lasted 1 s, after which the subject pressed “1” if the word had been seen in reading, and “0” if it had not. Both accuracy and reaction time were recorded. Following the design of Jones et al. (2012) and Nelson and Shiffrin (2013), each of the ten studied pseudowords was presented five times. The ten remaining unstudied pseudowords from the original set were used as foils, with each also presented five times, for a total of 100 trials. Unstudied and studied items were randomly intermixed, and no word was repeated sequentially. These design choices were carefully considered: Trial repetitions made it possible to determine the mean performance for each item, increasing the stability of the parameter estimate. Likewise, using a fixed (rather than random) foil set avoided possible differences in the distribution of targets and foils, which could have contributed to differential learning during test. Most importantly, these choices allowed for direct comparison with previous studies.

Following the PLDT, participants completed a semantic similarity judgment task. A pair of words was presented on screen, and subjects were asked to rate how similar the pair was in meaning on a scale from 1 to 7, with 1 being the least similar and 7 being the most similar. Pairs consisted of a studied pseudoword and a close associate of the pseudoword’s target meaning. Each of the ten studied items was paired with four close associates, yielding a total of 40 semantic similarity ratings.

### Model predictions

To establish what SDM predicts in these tasks, we trained the model on the same materials that our subjects received. For the PLDT, we compared the vector magnitude for each item following training over uniform passages against the magnitude following distinctive passages. A higher vector magnitude signals a greater strength in memory. As the top panel of Fig. 2 illustrates, the model predicts that items learned over diverse contexts should be represented more strongly in memory. Behaviorally, this suggests that subjects should be faster and more accurate at recognizing these items, as compared to those learned across uniform contexts.

It is worth noting that the use of pseudowords is essential to this prediction. In many tests of episodic memory, frequency of encounter does not map neatly onto memory strength, as it does in other lexical processing tasks, such as LDT and naming. Indeed, in a standard recognition task, with intentional encoding at study and a mixed list of high and low frequency



**Fig. 2** Predictions from the semantic distinctiveness model (SDM) after training on the same materials as our subjects. The top panel depicts the predicted memory strength for studied items. The bottom panel depicts predicted semantic similarity between studied items and target associates. Each panel compares predictions following low and high variability training contexts

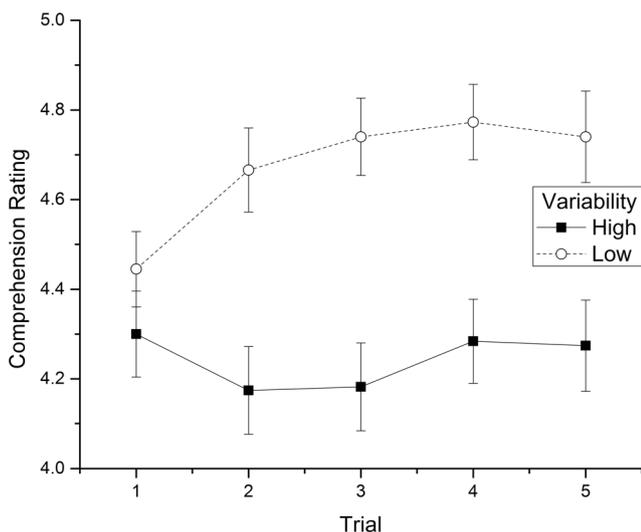
words, it is *low* frequency words that show a distinct processing advantage. However, the usual task design confounds a number of different contributing factors, including, for example, systematic differences in structural and semantic distinctiveness, differentiation in long-term memory, and contextual associativity (for discussion, see Nelson & Shiffrin, 2013). These confounds are far less of a concern in a task with randomly assigned pseudowords, where such properties can be manipulated or controlled through training materials. Unsurprisingly, episodic tasks that employ pseudowords report results that accord well with a strength-accrual account (e.g., Maddox & Estes, 1997). In the present study, pseudowords are the key element in mapping between the predictions of the SDM and the results of the PLDT.

To make predictions from the model for the semantic similarity rating task, we calculated the vector similarity (cosine) between each pseudoword and its target associate, and compared them across conditions. Representations of the associate words were obtained by training the model on a 200-k document Wikipedia corpus. Representations for pseudowords were constructed from the uniform or diverse paragraphs seen by subjects in training. Similarity between the pseudowords and the close associates was computed with a vector cosine. The model’s predictions are displayed in the bottom panel of Fig. 2. SDM predicts that items trained in uniform contexts should actually be *more* similar to their target associates than items trained in diverse contexts, as their high lexical overlap contributes to a more stable semantic representation. Given that a model based on frequency or a raw context count would predict no difference, this task is diagnostic in separating these models.

## Results

During the study phase of the experiment, subjects supplied comprehension ratings for each of the passages. A 2 (paragraph condition)  $\times$  5 (trial number) repeated measures ANOVA revealed a significant effect of paragraph diversity [ $F(1,86) = 110.26, p < 0.001$ ], a small effect of trial number [ $F(1,86) = 2.377, p = 0.05$ ], and a significant interaction [ $F(4, 344) = 4.565, p = 0.001$ ]. Figure 3 shows the average comprehension ratings for the low variability and high variability sets across the five passages. For the first passage, the ratings for the low and high variability passages are equivalent. However, for subsequent passages, the ratings for high variability passages are systematically lower, meaning they were rated as less comprehensible. By contrast, ratings increased over the low variability condition, indicating that participants' subjective comprehension of paragraphs within the same discourse topic grew as they gained more experience with that topic. In the high variability condition, the ratings are relatively stable by comparison, suggesting very little overlap in meaning of the different paragraphs across reading.

The pattern observed in comprehension judgments is mirrored in an examination of passage reading times. This post-hoc analysis, conducted at the request of a reviewer, relied on timing data reconstructed from experiment log files, which were recoverable for the majority of participants. It revealed that whereas low variability passages were studied for an average of 22.57 s, their high variability counterparts were studied 1.67 s longer (24.24 s), a significant difference [ $t(52) = 4.634, p < 0.001$ ]. Thus, variability appears to translate to longer study time. This is consistent with the model's prediction that when a target appears in an unfamiliar or unexpected context, more attentional resources will be allocated to



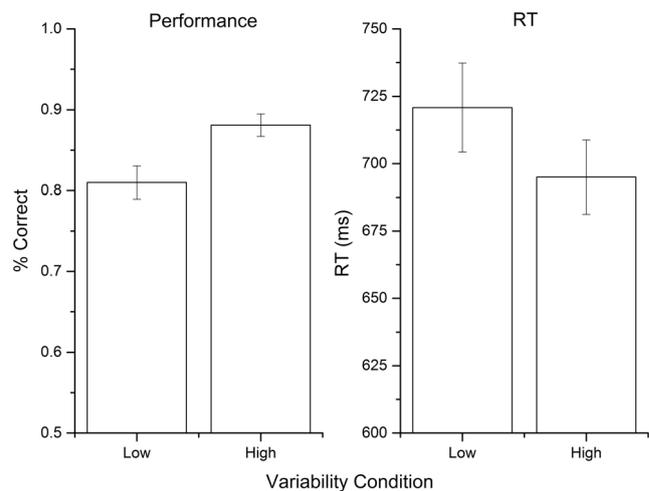
**Fig. 3** Comprehension ratings made on low and high variability passages across trials. Higher ratings indicate that the paragraphs were easier to comprehend

encoding it, leading to more efficient future identification latency for the target stimulus itself, but greater representational variance about its meaning.

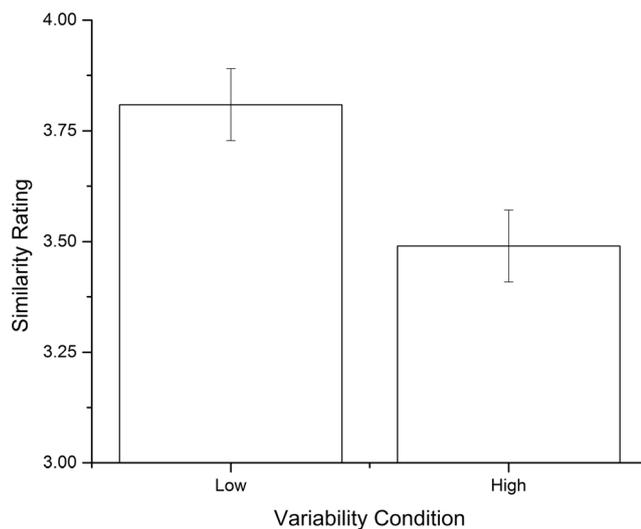
Turning then to the test phase, the SDM predicts that words that occur in more diverse semantic contexts should have a stronger representation in memory, making them easier to discriminate and faster to respond to. This prediction is supported by our PLDT results. In the PLDT, average accuracy was 83.9 % across conditions. As predicted, subjects were significantly more accurate at recognizing targets seen across highly variable contexts [ $t(86) = 3.561, p < 0.001$ ] (Fig. 4; left). Variability also appeared to support more rapid responding: Subjects were significantly faster at identifying words that appeared in high variability paragraphs [ $t(86) = 2.297, p < 0.05$ ] (Fig. 4; right), with a mean 26-ms advantage.

After completing the PLDT, subjects rated the semantic similarity of each pseudoword and four close associates of its target meaning. For our training materials, SDM predicts that items learned in uniform contexts should be rated as more similar to target associates than items seen across diverse contexts. In line with this prediction, subjects rated items trained on the low variability paragraphs as significantly more similar to their target associates [ $t(86) = 3.406, p = 0.001$ ] (Fig. 5).

These results reveal a dissociation between ease of processing and semantic representation early in learning. Subjects in our experiment appear to be more efficient at processing items trained over diverse contexts, recognizing those items more quickly and more accurately. At the same time, subjects appear to have better discriminated the meanings of items trained in redundant contexts, a finding supported both by their subjective comprehension ratings of the passages and by their increased similarity ratings in the semantic judgment task. These results closely mirror those of Hoffman and Woollams (2015), who found that for a non-randomly selected sample of



**Fig. 4** Performance and response time (RT) results from the pseudolexical decision task



**Fig. 5** Mean similarity ratings of studied items and target associates by passage training type

real words, contextual variability speeds lexical decision, while slowing semantic relatedness judgments.

## General discussion

Beyond early childhood, incidental learning from reading is one of the primary determinants of vocabulary growth (Nagy, Herman, & Anderson, 1985). In processing novel words, readers rely heavily on information available in the surrounding context, including both local distributional properties and broader world knowledge (McDonald & Ramscar, 2001). In this line of research, an open question is how the variability of the contexts in which a target is embedded influences its developing lexical and semantic representation.

In the experiment reported here, subjects were better at recognizing words after encountering them in highly variable contexts, but better at inferring their meanings after experiencing them across more stable semantic contexts, consistent with the predictions of the SDM model. The finding of more efficient lexical access for semantically diverse words is strongly coherent with previous results. However, this increased ease of processing actually led to poorer performance on a test of semantic similarity. That is, the semantic consistency of the low variability paragraphs allowed for a superior semantic representation to be formed, likely due to a greater ease of disambiguating the meaning of an unknown word in these contexts. This experiment points to the relativity of information in language learning: Different tasks are aided by different types of environmental information, and what may benefit one task may be harmful to performance on another.

We have thus far conceptualized this finding as a lexical-semantic effect, in which manipulations to the structure of the linguistic environment effected changes in the organization

and semantic representation of newly acquired words in the lexicon. However, this could also be seen as an episodic effect. Specifically, the shifts in semantic contexts in our experiment could be interpreted as an encoding variability manipulation (Bower, 1970), in which distinctive contexts lead to differential encoding, resulting in the observed differences in task performance. From this vantage, our experiment is one of pseudoword episodic recognition, rather than pseudoword lexical decision. Obviously, it is difficult to separate the contribution of language and memory on any task where words are used as stimuli (e.g., Johns & Jones, 2010; MacDonald & Christiansen, 2002). Nevertheless, it is a worthy question for future research as to how these systems interact in early learning.

While the SDM is capable of efficiently measuring contextual variability, and of making corresponding predictions about the effect that this should have on item recognition, it is only a representational model. However, its predictions align well with predictive accounts of language processing (e.g., Elman, 2009), in which speakers construct expectations about future linguistic input based on the current context. Words that are low in contextual variability will be better supported by consistent contextual cues, and thus should be weighted less strongly in memory, since they will be more predictable in context. Conversely, words that are high in contextual variability should be represented more strongly in the lexicon, since they are less associated with any given context, and thus lack contextual scaffolding. On this view, lexical access is a dynamic process, where both past experience with words and the current context combine to power retrieval.

**Acknowledgments** This work was supported by NSF grant BCS-1056744 and by IES R305A150546 to MNJ.

## References

- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814–823.
- Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*(4), 703–719.
- Balota, D. A., Cortese, M. J., Hutchinson, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.
- Bower, G. H. (1970). Organizational factors in memory. *Cognitive Psychology, 1*, 18–46.
- Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74*, 1–15.
- Church, K. W., & Gale, W. A. (1995). Poisson mixtures. *Natural Language Engineering, 1*(02), 163–190.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*(2), 452–478.

- Elman, J. L. (2009). On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. *Cognitive Science*, 33(4), 547–582.
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12(6), 627–635.
- Hills, T. T., Maouene, J., Riordan, B., & Smith, L. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, 63(3), 259–273.
- Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45(3), 718–730.
- Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance*, 1–19.
- Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, 41(6), 401–410.
- Johns, B. T., Dye, M., & Jones, M. N. (2014). The influence of contextual variability on word learning. In P. Bello, M. Guarani, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 242–247). Austin: Cognitive Science Society.
- 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(2), EL74–EL80.
- Johns, B. T., & Jones, M. N. (2008). Predicting word-naming and lexical decision times from a semantic space model. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 279–285). Austin: Cognitive Science Society.
- Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, 17, 662–672.
- Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, 66(2), 115–124.
- Jones, M. N., Willits, J. A., & Dennis, S. (2015). Models of semantic memory. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Oxford Handbook of Mathematical and Computational Psychology* (pp. 232–254). New York: OUP.
- 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.
- Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of Memory and Language*, 64(3), 249–255.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review*, 109, 35–54.
- Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 539–559.
- McDonald, S., & Ramscar, M. (2001). Testing the distributional hypothesis: The influence of context on judgments of semantic similarity. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society, CogSci'01* (pp. 611–616).
- Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, 76, 165–178.
- Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, 111(3), 721–756.
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233–253.
- Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, 120(2), 356–394.
- Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, 3(1), 1–17.
- Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 82–102.
- Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 760–766.
- Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 796–800.