Holographic stimulus representation and judgement of grammaticality in an exemplar model: Combining item and serial-order information

Randall K. Jamieson (randy_jamieson@umanitoba.ca)
Department of Psychology, University of Manitoba
Winnipeg, MB, R3T 2N2, Canada

D. J. K. Mewhort (mewhortd@queensu.ca)
Department of Psychology, Queen's University
Kingston, ON, K7L 3N6, Canada

Abstract

We examine representation assumptions for learning in the artificial grammar task. Strings of letters can be represented by first building vectors to represent individual letters and then concatenating the letter vectors into a vector of larger dimensionality. Although such a representation works well in selected examples of artificial-grammar learning, it fails in examples that depend on left-to-right serial information. We show that recursive convolution solves the problem by combining item and serial-order information in a stimulus item into a distributed data structure. We import the representations into an established model of human memory. The new scheme succeeds not only in applications that were successful using concatenation but also in applications that depend on left-to-right serial organization.

Keywords: Artificial grammar learning; Holographic representation; Exemplar model

Introduction

In an artificial-grammar learning (AGL) classification task, participants study strings of symbols. Following study, the participants are told that the studied items were constructed according to the rules of an artificial grammar and are invited to sort novel rule-based (grammatical) exemplars from novel rule-violating (ungrammatical) ones. Even though the participants are unable to describe the rules, they can discriminate the two classes of stimuli.

Initial accounts proposed that the participants abstracted the grammar and used that knowledge to judge the status of the exemplars (e.g., Reber, 1967, 1993). Later investigators argued that the participants judged grammaticality without reference to the grammar. To support the latter position, investigators identified several sources of information that discriminate the two classes of test strings. Brooks (1978) suggested that whole-item similarity between training and test strings is used to infer grammaticality. Perruchet and Pacteau (1990) argued that bigram overlap is used to infer grammaticality. Vokey and Brooks (1992) identified edit distance as a predictor, and Brooks and Vokey (1991) argued that patterns of repetition within a string are used to infer grammaticality. Knowlton and Squire (1996) identified associative chunk strength (ACS), and Johnstone and Shanks (1999) identified chunk novelty. Finally, Jamieson and Mewhort (2009a, 2010) showed that global similarity predicts performance in the task. Regression analyses designed to sort the various predictors have confirmed a role for all of them (e.g., Johnstone & Shanks, 1999). Factorial designs that have pitted predictors against one another have been unable to identify a single dominant predictor (e.g., Kinder & Lotz, 2009; Vokey & Brooks, 1992).

We think that many of the predictors (e.g., ACS, bigram over, etc) point to a common underlying factor, namely left-to-right serial structure. If so, the problem is not to determine which predictor dominates but, rather, to decide how subjects encode material so that they have access to the left-to-right serial structure in the exemplars.

In this paper, we explore an encoding mechanism that folds several orders of left-to-right serial structure in a string into a coherent and distributed data structure (i.e., single letters, bi-grams, trigrams, and whole strings). To begin,we describe the representation scheme. After, we show that the new representations predict judgement of grammaticality when used in an established theory of retrieval (Jamieson & Mewhort, 2009a, 2010).

Holographic representation in memory

Many investigators have proposed that light holography provides a mathematical basis for memory representation (Borsellino & Poggio, 1973; Gabor, 1968; Khan, 1998; Longuet-Higgins, 1968; Poggio, 1973). Murdock's (1982, 1983, 1997) TODAM is probably the best-known use of the idea in experimental psychology. In TODAM, stimulus associations are formed using linear convolution and associations are unpacked using correlation (deconvolution).

More recently, Jones and Mewhort (2007) used recursive circular convolution (Plate, 1995) to develop a self-organizing model of semantic memory (BEAGLE). BEAGLE captures judgements of semantic typicality, categorization, priming, and syntax from word order. BEAGLE's ability to handle so many phenomena of semantic memory is in itself impressive. However, from our perspective, BEAGLE's strength is that it shows how holographic representation can account for complex decision behaviour without adding control structures (e.g., learning and the application of rules). BEAGLE's success suggests that holographic stimulus representation should be explored in related models of learning and memory. The present work adapts BEAGLE's representation scheme to represent strings in the artificial grammar classification task.

Recursive circular convolution

Circular convolution is a mathematical operation that forms an associative representation, **z**, for two input vectors, **x** and **y**,

$$z_i = \sum_{j=0}^{n-1} x_{j \bmod n} \times y_{(i-j) \bmod n} \ , \qquad [1]$$

where i indexes the element in **z** and where n is the dimensionality of **z**, **x**, and **y**. Briefly, circular convolution forms the outer-product matrix—long used to represent associations in neural networks (e.g., Anderson, 1995)—and then collapses it into a vector (see Figure 1). Circular convolution is associative, commutative, and distributes over addition.

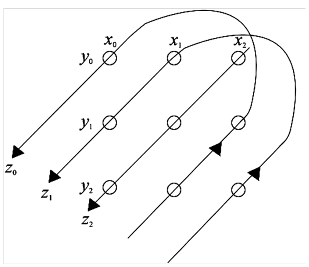

Figure 1. Collapsing an outer-product matrix with circular convolution, where **x** and **y** are the argument vectors and **z** represents the resulting compressed vector from collapsing the outer-product matrix. The values i and j represent the row and column indices, respectively, for an element in the outer-product matrix.

In the work that follows, we apply circular convolution recursively to encode a series, such as a sequence of letters. Consider the string *ABCD*. To represent *ABCD* as a series, first, generate a unique random vector for each of the individual letters in the string {**a**, **b**, **c**, **d**}. Next, apply circular convolution in a recursive fashion to bind the first letter to the second, the product of that binding to the third, and so on, until each letter has been folded into the representation. At this point, using **z** to represent the string *ABCD*, **z** = *ABCD* = ((**a** * **b**) * **c**) * **d**, where * denotes circular convolution. No matter the length of the string, **z** has the same dimensionality as the input (i.e., letter) vectors.

Why holographic representation?

In previous studies of AGL, we represented exemplars by concatenating letters. For example, a string *ABCD* was represented by concatenating the vectors for *A, B, C*, and *D* to form a single vector **a**//**b**//**c**//**d**, where // denotes concatenation. The scheme captured a swath of data from the artificial grammar task and from serial reaction time tasks (see Jamieson & Mewhort, 2009a, 2009b, 2010). Nevertheless, concatenated representations are problematic.

In models using vector representation, it is traditional to compute the similarity between **x** and **y**, using a vector cosine. Thus, with concatenated strings, similarity is computed by comparing information in corresponding serial positions of two strings (i.e., element-for-element). Because of the serial-position constraint, a model using the concatenated representation scheme treats the strings *ABCD* and *CDAB* as if they shared no overlapping features—a judgement that is at odds with data. In contrast, a holographic representation scheme distributes information throughout the vector so that each part of it contains some information about the whole. Thus, in difference to the concatenation scheme, the cosine calculation compares all parts of **x** (i.e., *ABCD*) and **y** (i.e., *CDAB*) simultaneously and, thereby, acknowledges similarity between *ABCD* and *CDAB*. Because participants appreciate the similarity between *ABCD* and *CDAB*, the holographic scheme is preferred.

Critically, holographic stimulus representation finesses the problem of encoding serial structure. Importantly, it does so without requiring a change in the similarity calculation or other aspects of retrieval. This occurs because a representation of *ABCD* that is formed using recursive circular convolution superimposes overlapping orders of serial structure onto a single distributed structure. Because different orders of serial information about a string are superimposed in a single representation, a standard cosine of two vectors supports parallel comparison of multiple orders of serial structure. The question we pose, then, is if we import the holographic representations into an established model of retrieval, will the previously successful model still work; that is, can we still explain peoples' judgements in the artificial grammar task?

Minerva 2

Minerva 2 is an established model of retrieval (Hintzman, 1986, 1988). When a participant studies an item, an event is encoded to memory as a unique trace.

In Minerva 2, a stimulus is represented by a vector of n elements; each element takes values: +1 or -1. To represent stimuli in the artificial grammar task, we first, generate a unique random vector for each of the letters in the English alphabet and then apply recursive circular convolution to those vectors to represent a string of letters. Thus, a string *TXXV* is represented by a trace: ((**t** * **x**) * **x**) * **v**.

Memory is a matrix, **M**. Encoding an event involves copying its corresponding vector representation to a new row in the memory matrix. Encoding is sometimes imperfect. Imperfect encoding is implemented by setting some vector elements to zero (indicating that the element is indeterminate or unknown). A parameter, L, controls the probability with which an element is stored. As L increases, encoding quality improves.

All retrieval is cued. When a retrieval cue is presented, it activates each trace in memory in proportion to its similarity to the cue. The activated traces are aggregated into a composite called the *echo*; the contribution of each trace to the echo is based on its activation.

The similarity of trace, i, to the probe, P, is computed using a vector cosine, i.e.,

$$S_i = \frac{\sum_{j=1}^{n} P_j \times M_{ij}}{\sqrt{\sum_{j=1}^{n} P_j^2}\sqrt{\sum_{j=1}^{n} M_{ij}^2}}, \qquad [2]$$

where P_j is the value of the j^{th} feature in the probe, M_{ij} is the value of j^{th} feature of the i^{th} row in memory. Like the Pearson r, the similarity of the i^{th} item to the probe, S_i, is scaled to the interval [-1, +1]. Similarity equals +1 when the trace is identical to the probe, 0 when the trace is orthogonal to the probe, and -1 when the trace is opposite to the probe.

The i^{th} trace's activation, A_i, is the cube of its similarity to the probe, i.e.,

$$A_i = S_i^3. \qquad [3]$$

The activation function exaggerates differences in similarity between a probe and items in memory by attenuating activation of exemplars that are not highly similar to the probe. This allows traces most similar to the probe to dominate the information retrieved. Note that the exponent in the activation function preserves the sign of the argument, S_i.

The information retrieved by a probe is a vector, **c**, called the echo. The echo is computed by weighting each of the i = 1 ... m traces in memory by their activations and, then, summing all m traces into a single vector,

$$c_j = \sum_{i=1}^{m} A_i \times M_{ij}. \qquad [4]$$

The information in the echo is converted to decision variable called echo intensity, I, by computing the cosine similarity (see Equation 2) of the echo and probe. In the context of the artificial grammar task (i.e., classification), echo intensity is a proxy for judgement of grammaticality.

In the remainder of this paper we apply the model to data from the judgment of grammaticality task.

Evaluating the model

The judgement of grammaticality task was introduced by Reber (1967). In his experiment, participants memorized grammatical exemplars. After, they judged the grammatical status of novel test probes. Reber's subjects discriminated novel grammatical from novel ungrammatical test probes, but they could not articulate the rules of the grammar.

We have shown previously that Minerva 2 captures discrimination of grammatical from ungrammatical test probes in Reber's (1967) task, without reference to grammatical rules (Jamieson & Mewhort, 2009a, 2010). But we used concatenated stimulus vectors in that work. In the simulations that follow, we retest the model using the holographic rather than concatenated stimulus vectors.

To simulate Reber's (1967) task we began by representing his stimuli in our model.[1] First, we constructed a unique 100-element vector to represent each letter used to construct letter strings: $\{T, V, P, X, S\}$. Second, we generated a vector for each training and test string using recursive circular convolution. Third, we filled successive rows on the memory matrix with the training vectors. Fourth, we introduced moderate data-degradation to the items in memory, i.e., L = 0.7. Finally, we calculated the mean echo intensity for each of the 24 grammatical and 24 ungrammatical test strings.

The new model successfully discriminated grammatical from ungrammatical test items. The mean echo intensity for the 24 grammatical test strings was .57 (SD = .03); the corresponding value for the 24 ungrammatical test strings was .49 (SD = .02), $t(48) = 2.15$, $p < .05$.

In other simulations, we varied the integrity of data in memory (e.g., Jamieson, Holmes, & Mewhort, in press). As shown in Figure 2, the magnitude of the difference in mean echo intensity for grammatical and ungrammatical test strings (i.e., the model's discrimination of grammatical and ungrammatical items) grew as a function of L.

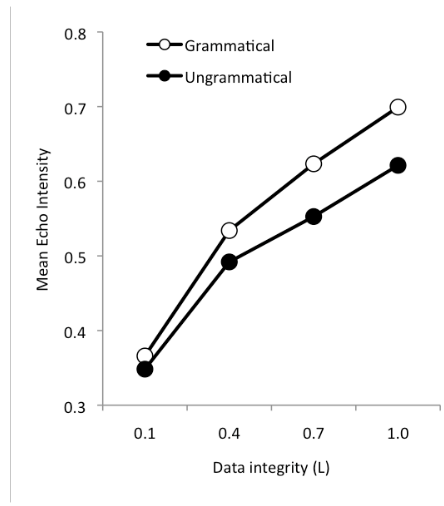

Figure 2. Mean echo intensity for grammatical and ungrammatical test strings as a function of data integrity in memory, L.

The simulation illustrates several points. First, it shows that the distributed stimulus representations generated using recursive circular convolution support discrimination of grammatical from ungrammatical test items. Second, because the model discriminated the two classes of stimuli without reference to grammatical rules, the simulation serves as an existence proof that grammatical strings can be

[1] Reber did not list the specific study and test items that he used in his original paper. He did, however, provide a list of representative strings from the same grammar in another source (Reber, 1993, p. 36). We took our strings from there.

discriminated from ungrammatical test strings without knowledge of the grammatical rules. Thirdly, the simulation shows that we can import holographic stimulus representations into Minerva 2 without a deleterious impact on the effects that the model captures using concatenated vectors (see Jamieson & Mewhort, 2009a, 2010).

Next, we test the new representation scheme by applying it to data collected by Kinder and Lotz (2009). Their data provide a more detailed challenge.

Kinder and Lotz (2009)

Kinder and Lotz (2009) engineered an artificial grammar to distinguish stimulus properties thought to predict judgements of grammaticality. They used the grammar to construct a list of 12 training items and 48 test items. The test items were of four different types. Type 1 and Type 2 items were ungrammatical; Type 3 and Type 4 items were grammatical. Type 1 test items violated both positional and sequential rules of the grammar; Type 2 items violated only sequential rules (i.e., the strings included at least one illegal bigram but all letters were in legal serial positions). Type 3 and Type 4 items obeyed positional and sequential rules of the grammar; but, Type 4 items had the additional property of being very similar to a specific training exemplar. Accordingly, if participants endorse Type 2 over Type 1 items, they must appreciate the positional dependencies of letters in the training set. If participants endorse Type 3 over Type 2 items, they must appreciate the difference between studied and unstudied chunks (i.e., bigrams and trigrams). If they endorse Type 4 over Type 3 items, they must appreciate whole-item similarity between training and test strings.

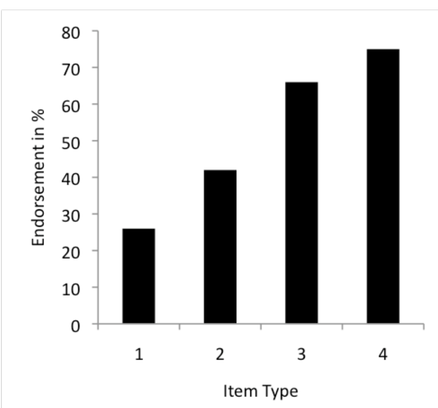

Figure 3. Empirical: Percentage of items endorsed as grammatical in Kinder and Lotz's (2009) Experiment 2.

Kinder and Lotz's (2009) results are reproduced in Figure 3. First, participants endorsed Type 2 over Type 1 items indicating they were sensitive to the positions of individual letters in the training strings. Second, participants' endorsed Type 3 over Type 2 items, indicating they were sensitive to test strings' inclusion/exclusion of studied and unstudied bigrams. Finally, participants endorsed Type 4 over Type 3 items, indicating they were sensitive to whole-item similarity between training and test strings.

The pattern of results demonstrates that judgement of grammaticality is influenced concurrently by the positions of single letters in a string, by knowledge of small chunks (i.e., knowledge of bigrams and trigrams), and by knowledge of larger chunks (i.e., whole training strings). To claim a model as a competent account of decision in the judgement of grammaticality task, the model must accommodate concurrent sensitivity to the three sources of information.

Simulation of Kinder and Lotz (2009; Exp 2)

Kinder and Lotz's (2009) data provide a principled challenge to test the idea that holographic stimulus representation allows multiple orders of serial-structure to exert a concurrent influence on judgements of grammaticality. Hence, we tested our model using Kinder and Lotz's (2009) materials.[2] The simulation was otherwise the same as before.

The results of the simulation are presented in Figure 4; the means were computed across 50 independent replications of the procedure. We treat mean echo intensity as a proxy for mean judgement of grammaticality.

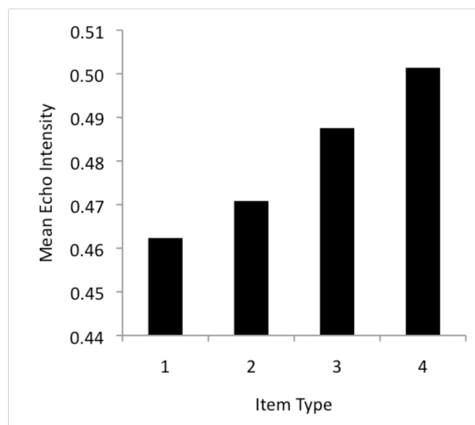

Figure 4. Simulation: Mean echo intensity for the four item types in Kinder and Lotz's (2009) Experiment 2.

As shown, the model reproduced the pattern of results from Kinder and Lotz's (2009) experiment. Firstly, mean echo intensity for Type 2 items was greater than for Type 1 items indicating that the model was sensitive to positional dependencies of individual letters in the training strings. Secondly, echo intensity for Type 3 items was greater than for Type 2 items indicating that the model was sensitive to bigram and trigram structure in the stimuli. Finally, echo intensity for Type 4 items was greater than for the Type 3

[2] A complete listing of Kinder and Lotz's (2009) materials is presented in their Appendix B. The simulations were identical for the two sets; a testament to their care at stimulus design.

items indicating that the model was sensitive to larger chunks of letters, possibly whole strings.

Importing a scheme for holographic stimulus representation into a Minerva 2-based account of retrieval allows the model to capture additional details of performance in the artificial grammar task. Minerva 2 now captures trends that previously required a very different kind of computational account (e.g., the SRN, see Elman, 1990).

General Discussion

Judgements of grammaticality reflect a concurrent consideration of discriminative cues (e.g., Johnstone & Shanks, 1999). To accommodate that fact, we developed a new kind of stimulus representation based on recursive circular convolution. The new representation folds information about several cues into a distributed data structure. More importantly, the holographic representation scheme supports parallel comparison of features in a string, unconstrained by serial position alignment. Using the holographic representations in a model of human memory captures judgement of grammaticality.

In previous work, Jamieson and Mewhort (2009a, 2010) demonstrated that judgment of grammaticality can be understood using Minerva 2—an exemplar model of memory. In that work, exemplars were represented by concatenating individual letter vectors. Judgement of grammaticality reflected a test probe's global similarity to the studied exemplars. The representation scheme worked because it preserved the spatial structure of the stimulus (i.e., letters from left-to-right). However, the account neglected to include information about left-right sequential properties of the exemplars—information that subjects notice during study. Because the model did not acknowledge sequential structure in stimuli, it incorrectly computed similarity between two exemplars based on bigram overlap; a factor measured by associative chunk strength.

The holographic stimulus representations developed here finesse the problem associated with the earlier scheme by folding information about serial-structure into the representation of a string. By using the holographic representations, the model now captures judgements that reflect serial structure (e.g., participants' appreciation of chunk overlap). Despite changes to the representation scheme in the model, we have not changed the model's account of retrieval and so we retain our previous conclusion: Judgement of grammaticality can be captured without an implicit rule-induction process that abstracts and applies grammatical information.

Kinder (2000; Kinder & Lotz, 2009) and others (e.g., Cleeremans, Servan-Schreiber, & McClelland, 1989) have argued for a Simple Recurrent Network (SRN) account of artificial grammar learning. The SRN accomplishes judgement of grammaticality by learning the sequential structure in a set of training sequences and then applying that knowledge to predict sequential regularities in test items. When the SRN can predict the sequential structure of a test string, it judges the test string as grammatical (see Reber, 2002, for an analysis of the approach; see Vokey & Higham, 2004, for model comparison of the SRN and a related instance-based model). Cleeremans et al. (1989) showed the SRN develops a veridical representation of the grammar used to generate the training strings. By contrast, our account treats judgement of grammaticality as an episodic memory task. At study, the model encodes information about individual exemplars, including serial structure. At test, the model judges a test strings' grammaticality by its global similarity to the exemplars in memory. The two classes of model (Minerva 2 and the SRN) offer very different explanations of the cognitive processes that underlie judgement of grammaticality. So, which approach is to be preferred? We think the answer should be based on the nature of the experimental problem.

In the training phase of a standard artificial grammar experiment, participants are asked to memorize exemplars. At test, they are given the problem of inferring the grammaticality of test probes. Of course, people *can* learn sequential structure in stimuli instructions. But they do not have to learn it: the task does not cue them to do so. In our view, although learning sequential structure in a set of exemplars provides a possible mechanism, for the judgement of grammaticality task, it implies compulsory learning of sequential regularities even though that action is neither implied by nor cued by the task. Unlike the SRN, Minerva 2 assumes people notice sequential characteristics of each exemplar, but they do not learn the regularities in the set of exemplars. Moreover, because our account treats judgement of grammaticality as a retrospective judgement, it is not necessary to justify or to describe prospective abstraction of structure in the training set.

In developing our holographic representation scheme, we have been careful to avoid altering our model's assumptions about retrieval. In both our original and our present accounts, we assumed a perceptual system loads memory with what the subjects notice about each of the studied exemplars. Judgment of grammaticality reflects the global similarity of a test probe to training items. The difference in our new account is that the new model assumes that subjects notice more about the order of the symbols than the old model assumed; a claim echoed in post-experimental interviews with our subjects. At a broader level, our solution honours an insight from Simon's (1969) parable of the ant. Simon noted that an ant's path on a beach may be complex and difficult to describe. But, the complexity of the path may be driven by complexity in the beach rather than complexity in the ant. Simon used the parable to goad theorists into considering explanations for a behaviour based on the complexity of the environment before assuming that the behaviour reflects complex psychological mechanisms. Here, we have followed Simon's advice. Peoples' behaviour in the artificial grammar task appears complex and difficult to describe. However, the complexity is in the materials, not in the subjects. Judgement of grammaticality reflects the storage and retrieval of studied exemplars.

Acknowledgments

R. K. Jamieson, Psychology, University of Manitoba; D. J. K. Mewhort, Psychology, Queen's University. The research was supported by grants to both authors from the Natural Sciences and Engineering Research Council of Canada. Mail correspondence to R. K. Jamieson, Department of Psychology, University of Manitoba, Winnipeg, MB, Canada, R3T 2N2. Electronic mail to: randy_jamieson@umanitoba.ca.

References

Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.

Borsellino, A., & Poggio, T. (1973). Convolution and correlation algebra. *Kybernetik*, *122*, 113-122.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, *120*, 316-320.

Cleeremans, A., Servan-Schreiber, D., & McClelland, J. L. (1989). Finite state automata and simple recurrent networks. *Neural Computation*, *1*, 372-381.

Dienes, Z. (1993). Computational models of implicit learning. In D. C. Berry & Z. Dienes (Eds.), *Implicit learning: Theoretical and empirical issues* (pp. 81-112). Hove, UK: Lawrence Erlbaum Associates.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Gabor, D. (1968). Improved holographic model of temporal recall. *Nature*, *217*, 1288-1289.

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

Hintzman, D. L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528-551.

Jamieson, R. K., Holmes, S., & Mewhort, D. J. K. (2010). Global similarity predicts dissociation of classification and recognition: Evidence questioning the implicit/explicit learning distinction in amnesia. *Journal of Experimental Psychology: Learning Memory and Cognition*, in press.

Jamieson, R. K., & Mewhort, D. J. K. (2010). Applying an exemplar model to the artificial grammar task: String-completion and performance on individual items. *Quarterly Journal of Experimental Psychology*, *63*, 1014-1039 .

Jamieson, R. K., & Mewhort, D. J. K. (2009a). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology*, *62*, 550-575.

Jamieson, R. J., & Mewhort, D. J. K. (2009b) Applying an exemplar model to the serial reaction-time task: Anticipating from experience. *Quarterly Journal of Experimental Psychology*, *62*, 1757-1783.

Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 524-531.

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

Khan, J. I. (1998). Characteristics of multidimensional holographic associative memory in retrieval with dynamically localizable attention. *IEEE Transactions on Neural Networks*, *9*, 389-406.

Kinder, A. (2000). The knowledge acquired during artificial grammar learning: Testing the predictions of the two connectionist models. *Psychological Research*, *63*, 95-105.

Kinder, A., & Lotz, A. (2009). Connectionist models of artificial grammar learning: What type of knowledge is acquired? *Psychological Research*, *73*, 659-673.

Knowlton, B. J. & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22,* 169-181.

Longuet-Higgins, H. C. (1968). Holographic model of temporal recall. *Nature*, *217*, 104.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609-626.

Murdock, B. B. (1983). A distributed model for serial-order information. *Psychological Review*, *90*, 316-338.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, *104*, 839-862.

Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, *6*, 623-641.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*, 264-275.

Poggio, T. (1973). On holographic models of memory. *Kybernetik*, *12*, 237-238.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior*, *6*, 855-863.

Reber, P. J. (2002). Attempting to model dissociations of memory. Trends in Cognitive Sciences, 6, 192-194.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Simon, H. A. (1969). The psychology of thinking: Embedding artifice in nature (Chapter 2). In *Sciences of the artificial (pp. 23-26)*. MIT Press.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1504-1510.

Vokey, J. R., & Higham, P. A. (2004). Opposition logic and neural network models in artificial grammar learning. *Consciousness and Cognition*, *13*, 565-578.