Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale
2012, Vol. 66, No. 2, 98–105

1196-1961/12/$12.00 DOI: 10.1037/a0027023

An Exemplar Model of Performance in the Artificial Grammar Task: Holographic Representation

Randall K. Jamieson and Brian R. Hauri
University of Manitoba

We apply a multitrace model of memory to explain performance in the artificial grammar task. The model blends the convolution method for representation from Jones and Mewhort's BEAGLE model (Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114,* 1–37) of semantic memory with the multitrace storage and retrieval model from Hintzman's MINERVA 2 model (Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428) of episodic memory. We report an artificial grammar experiment, and we fit the model to those data at the level of individual items. We argue that performance in the artificial grammar task is best understood as a process of retrospective inference from memory.

Keywords: exemplar model, holographic reduced representation, artificial grammar, MINERVA 2

In an artificial grammar task, participants study letter strings constructed according to the rules of an artificial grammar. Afterward, they discriminate grammatical rule bound from ungrammatical rule violating test strings. Typically, people can discriminate the two classes of test items, but they cannot articulate the grammar.

Three kinds of theories have been proposed to explain peoples' performance. *Abstractionist theories* propose that participants internalize the grammar and, at test, endorse strings that match it. When participants cannot articulate the grammar, grammatical knowledge is diagnosed as implicit (e.g., Mathews et al., 1989; Reber, 1967). *Statistical theories* propose that participants learn stimulus regularities (e.g., bigram frequencies) and, at test, endorse strings that exhibit the regularities. The statistical theories divide on whether participants' knowledge of statistical regularities is implicit (e.g., Dulany, Carlson, & Dewey, 1984; Knowlton & Squire, 1996; Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). *Exemplar theories* propose that participants store the training exemplars and, at test, endorse strings that remind them of the studied exemplars. Because no implicit knowledge of the grammar is assumed, the discrepancy between performance and awareness of the grammar is irrelevant (e.g., Brooks, 1978; Higham, 1997; Jamieson, Holmes, & Mewort, 2010; Jamieson & Mewhort, 2009, 2010; Nosofsky & Zaki, 1998; Pothos & Bailey, 2000; Vokey & Brooks, 1992; Wright & Whittlesea, 1998).

Jamieson and Mewhort (2009, 2010) formalized the exemplar account using Hintzman's (1986) multitrace model of memory. According to the model, participants store studied exemplars in memory. At test, a probe retrieves an aggregate of the stored traces. Judgment of grammaticality is based on the match between the probe and the retrieved aggregate. The model tracks judgment of grammaticality and serves as proof of concept that performance in the artificial grammar task can be understood as an example of exemplar-based inference.

Despite the demonstrated capabilities of Jamieson and Mewhort's (2009, 2010) retrieval model—close fits to data from experiments—it was limited by the constraints imposed by a naïve model for representation. To illustrate, a letter string *ABCDEF* is represented by a vector composed of six successive subfields **a**//**b**//**c**//**d**//**e**//**f**, where // indicates concatenation. Although the scheme maintained the spatial structure of the letter string, it was wrong. It was wrong because it contradicted data that show people do not remember a stimulus as a literal spatial object, but rather remember it as encoded (e.g., Higham, 1997; Vokey & Brooks, 1992; Wright & Whittlesea, 1998).

The difference between a stimulus as presented and a stimulus as encoded has been long appreciated. For example, Miller (1958), who published the seminal study with the artificial grammar task, argued that an account of *recoding* was central to explain the recall benefit for grammatical over ungrammatical sequences. For example, when presented with a string *RGBYRGBY*, a participant can recode the sequence as *RGBY twice*—a recoding strategy that halves the memory load of the original stimulus and thereby eases the difficulty of recall.

Jamieson and Mewhort (2005) built on Miller's (1958) position and showed that grammatical sequences, on average, afford more efficient recoding than do random sequences. Thus, an account of recall that considers recoding can explain the well-documented recall benefit for grammatical over random sequences.

Recoding is also observed in the judgment of grammaticality task. For example, when presented with a letter string such as *RGBYBY*, people tend to encode the string as a collection of subunits, rather than as a whole (e.g., *RG*, *BY*, and *BY*, or *RGB* and

Randall K. Jamieson and Brian R. Hauri, Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada.

Randall K. Jamieson is indebted to D.J.K.M. The research was supported by a Discovery Grant to Randall K. Jamieson from The National Sciences and Engineering Research Council of Canada.

Correspondence concerning this article should be addressed to Randall K. Jamieson, Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2. E-mail: randy_jamieson@umanitoba.ca

YBY). Accordingly, it has been proposed that judgment of grammaticality reflects a comparison between the subunits in a test string with subunits in the training set (e.g., Perruchet & Gallego, 1997; Perruchet & Pacteau, 1990).

To accommodate the empirical facts, Jamieson and Mewhort (2011) proposed a representation scheme that allowed for chunk-based encoding and that was based on the mathematics of holography. Their method relied on Gabor's (1968, 1969) insights that a representation scheme based in vector convolution mimics a hologram (see also Longuet-Higgins, 1968; Metcalfe-Eich, 1982; Murdock, 1982, 1995; Poggio, 1973). To implement holographic representation, Jamieson and Mewhort used Jones and Mewhort's (2007) noncommutative circular convolution (Plate, 1995). In the model, each letter was represented by a unique vector. Each letter string was represented by sampling subsequences from it. The representation of an exemplar was formed by summing the subunit vectors. Combined with the holographic representation scheme, Jamieson and Mewhort's model accommodated data from three experiments that their retrieval model combined with concatenated representation did not. They argued that, based on the difference, holographic representation is a better method for capturing what people notice and how they represent training and test strings in memory. Jones and Mewhort used the operation to encode subsequences from linguistic input and, thereby, induce the semantic and grammatical structure in a linguistic corpus. Jones and his colleagues have extended the method to an examination of how one might encode the details of word form (Cox, Kachergis, Recchia, & Jones, 2011; see also Hannagan, Dupoux, & Christophe, 2011). Others have used the method to model the behavior of spiking neurons (Eliasmith, 2004).

Whereas Jamieson and Mewhort (2011) established the method for holographic representation within a standard artificial grammar task, their analysis was brief. Namely, they fit data to just a few classes of items in just a few experiments. Thus, the generality of their argument remains an open question.

Dienes (1992) has argued that a competent model ought to do more than fit performance for classes of items. Namely, a competent model ought to also fit the pattern of performance over individual test items. Dienes' dictum is important because it calls for a model of performance to predict decisions for stimuli rather than decisions for stimulus properties. It is also important because it forces an increasingly complete account of performance. In the work that follows, we take up Dienes' challenge and test the ability of Jamieson and Mewhort's (2011) model to predict performance on individual items.

Experiment

We conducted a standard artificial grammar task using materials like those from Reber's (1967) classic study. In a training phase, participants studied 20 training strings. At test, participants rated the grammaticality of test strings using a slider. The slider was marked *Rule Violating* at its extreme left, *Unsure* at its midpoint, and *Rule Conforming* at its extreme right. For the analysis of responses, the extreme left of the slider corresponded to a rating of −100 and the extreme right corresponded to a rating of +100.

We anticipated a usual result: Participants will rate grammatical test items as more grammatical than ungrammatical test items, but they will claim ignorance of the grammar.

Method

Participants

A total of 52 students from the University of Manitoba undergraduate participant pool took part in the study. All participants reported normal or corrected-to-normal vision.

Apparatus

The experiment was administered on personal computers (PCs). Each PC was equipped with a 21.5-in. wide-screen monitor, a standard keyboard, and a standard mouse. Participants responded using the mouse to click on words displayed on the monitor and with the keyboard to report the rules of the grammar.

Materials

The stimulus materials were taken directly from Reber (1993, p. 36). The stimulus set included 20 grammatical training strings, 25 grammatical test strings, and 25 ungrammatical test strings. Six of the 25 grammatical test strings appeared in both the training and test lists; the remaining 19 grammatical test strings did not. A string's grammatical status was defined by its conformity to the grammar in Figure 1. Any string that can be generated with the grammar is "grammatical." Any string that cannot be produced with the grammar is "ungrammatical." We used Reber's materials instead of generating our own to preempt criticism that we used special items or that we engineered our stimuli to maximize model fit. By Reber's own definition, our materials are representative of materials used in the artificial grammar task.

Procedure

Participants were tested in groups of three to seven.

After participants were seated at different computers, they were told that they would be shown strings of letters and that they

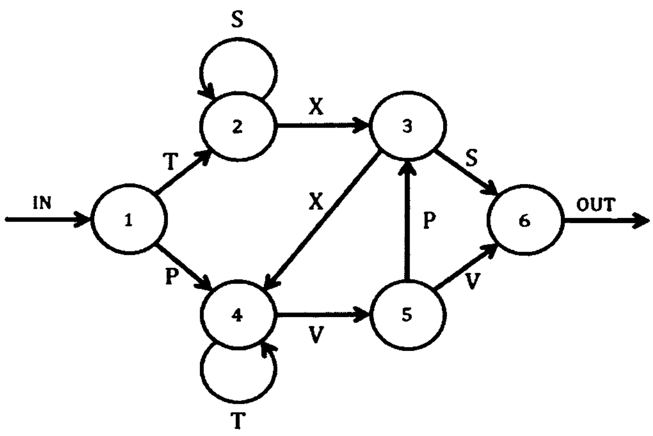

Figure 1. Finite state grammar used to construct the materials. A grammatical stimulus is generated by starting at the leftmost node marked 1 and following the paths through the diagram (indicated by arrows) until reaching the rightmost node marked 6. When a path is taken the associated letter is added to the end of the string. For example, moving from nodes 1 to 2, 2 to 3, and 3 to 6 produces the string *TXS*.

should memorize them for a test. The participant initiated the training phase by clicking on the message "Start" that was displayed at the center of the computer screen. After the participant had clicked on the message, the screen was cleared for 750 ms. Immediately thereafter, the first training string was displayed at the center of the screen for 6 s. Next, the screen was cleared, and after 750 ms, the next string was displayed. The cycle repeated until all of the training strings were presented.

Following the training phase, the participants were told that all of the 20 strings that they had studied had been constructed according to rules of an artificial grammar and that it was now their task to rate the grammaticality of novel test strings. On each test trial, a test string was presented at the center of the screen. A line approximately 5 cm in length was displayed approximately 3 cm below the string: The line contained a slider positioned at its center. The phrases "Rule Violating," "Unsure," and "Rule Conforming" were displayed at the left, center, and right of the line, respectively. A button labeled "OK" was centered approximately 2 cm below the line.

At the start of each test trial, the slider was positioned at the center (neutral point) of the line. To rate the grammaticality of the test string, the participant used the computer mouse to position the slider on the line and then clicked on the word OK. Immediately thereafter, the screen was cleared and, 1 s later, the next string was displayed. If the participant did not move the slider from the neutral position before clicking on the word OK, a message instructed the participant that he or she must move the slider from the neutral position in order to complete the trial. The cycle continued until all of the test strings had been presented.

Although participants were not shown numbers on the slider, the left extreme corresponded to a rating of -100, the right extreme corresponded to a rating of $+100$, and the midpoint of the line corresponded to a rating of 0.

After the test strings had been presented, a text editor appeared on the computer screen. A message above the text editor invited the participant to describe the rules that were used to construct the training strings. Participants provided their rule reports by using the keyboard to type their responses into the text editor. When they were finished, they clicked on a button marked OK.

Results and Discussion

Participants discriminated the grammatical ($M = 18.97$, $SE = 2.63$) from the ungrammatical ($M = -9.72$, $SE = 2.39$) test strings, $t(51) = 11.50$, $d = 1.59$, $p < .05$. However, none of the participants articulated the rules of the grammar; indeed, most declined to guess when pressed. The results are consistent with standard results of artificial grammar experiments and with Reber's (1967) classic study using like materials.[1]

Table 1 shows the mean grammaticality ratings for all 50 of the test strings; items 1 through 25 in Table 1 are grammatical and items 26 through 50 are ungrammatical. Bolded items were presented at both training and test. The table indicates items with mean ratings that differed reliably from zero, $\alpha = .05$.

We note five aspects of the data in Table 1. First, there is a spread in the ratings for items both between and within categories, suggesting that participants judged test items based on factors other than grammatical status. Second, ratings for only 31 of the 50 test items were reliably different from zero, $\alpha = .05$, showing that whereas participants agreed on the grammatical status of some items, they disagreed on the grammatical status of others. Third, although participants rated most test strings in agreement with their true grammatical statuses, participants rated four test strings in strong disagreement with their true grammatical statuses: items 11, 17, 40, and 48. Fourth, although participants rated all six test strings that were presented at both training and test positively, they did not rate the six studied items abnormally positively—a result that is consistent with data published by Kinder and Shanks (2001; see the bolded items in Table 1). Finally, we failed to detect that the means in Table 1 were explainable by a simple decision rule or idiosyncratic distortion (e.g., string length, inclusion/exclusion of a single bigram or trigram such as *TV*, inclusion/exclusion of single-letter recursions, or legality of first letter).

The data in Table 1 provide a valid database for evaluating Jamieson and Mewhort's (2011) holographic model of performance in the artificial grammar task. First, the data were collected using representative materials from a representative grammar (Reber, 1993). Second, the data were collected using a simple and theory-neutral study procedure (i.e., memorization). Third, performance averaged over participants and within stimulus classes was consistent with published data: participants discriminated grammatical from ungrammatical items but could not articulate the grammar. Finally, the data are sufficiently resolved to enable an item level analysis of performance.

We now outline Jamieson and Mewhort's (2011) model and then apply it to the design and materials from Experiment 1. If the model is competent, it will discriminate grammatical from ungrammatical test items as well as reproduce the profile of mean ratings from Table 1.

Holographic Exemplar Model

The holographic exemplar model (HEM) merges the holographic representation model from Jones and Mewhort's (2007) BEAGLE with the storage and retrieval model in Hintzman's (1986) MINERVA 2.

Representation

In the HEM, a letter is an n-dimensional vector. Each element in a letter vector takes a value sampled from a normal distribution with mean zero and variance $1/n$. Letter strings are also represented as n-dimensional vectors, but they are formed by applying noncommutative circular convolution to the letter vectors.

Circular convolution is a vector operation that encodes an association between two vectors, $\mathbf{x}$ and $\mathbf{y}$, to a new vector, $\mathbf{z}$:

$$z_i = \sum_{j=0}^{n-1} x_{j \bmod n} \times y_{(i-j) \bmod n} \{for\ i = 0 \ldots n-1\}, \quad (1)$$

[1] The overall pattern of performance in our experiment matched that in Reber's (1967). However Reber's subjects discriminated grammatical from ungrammatical test strings much better than did our subjects. The difference most likely reflects the fact that we presented each training string once (and only briefly), whereas Reber presented items to subjects as many times as they needed (and for long presentation) until they could report the set. Another problem for direct comparison is that our materials are representative of Reber's materials—they are not the exact materials that Reber used.

Table 1
Mean Ratings of Grammaticality for Test Items

Item number	Item	Grammaticality ratings M	SE	$p < .05$
1	PTTTTVPS	−1.94	9.82	
2	**PVV**	27.92	10.34	a
3	TXXTTVPS	44.17	7.18	a
4	TSXXTTVV	32.25	7.82	a
5	TSSXXTVV	38.80	8.26	a
6	**TXS**	29.12	10.10	a
7	TSXXVPS	41.85	7.44	a
8	PTTTVV	−3.69	8.26	
9	PTTVPS	36.94	7.94	a
10	TXXTTVV	21.15	8.59	a
11	PTTTTTVV	−25.67	9.54	a
12	TSSXS	27.76	7.76	a
13	PVPXTTVV	8.31	8.57	
14	PTTVPXVV	25.35	7.13	a
15	**TSXXTVPS**	31.28	8.78	a
16	TXXTVV	41.31	7.85	a
17	TSSSSXS	−28.39	9.55	a
18	TPVV	−10.88	8.02	
19	**PVPXTVPS**	25.98	9.10	a
20	**PVPXVV**	25.67	8.09	a
21	PTVPXVPS	8.13	8.41	
22	TXXVV	19.56	9.16	a
23	**TSSXXVPS**	38.65	8.15	a
24	TSXXVV	32.37	7.34	a
25	PVPS	4.17	9.43	
26	PTTTVPVS	3.12	8.30	
27	PVTVV	6.65	8.23	
28	TSSXXVSS	12.40	8.16	
29	TTVV	−2.17	9.77	
30	PTTPS	−1.42	8.18	
31	PVXPVXPX	−44.38	7.73	a
32	XXSVT	−38.20	6.57	a
33	TXXVX	−9.08	8.11	
34	PTTTVT	−2.29	9.01	
35	TXV	−19.08	8.87	a
36	PSXS	−6.23	8.42	
37	PTVPPPS	−18.37	7.65	a
38	TXVPS	3.08	8.41	
39	SVPXTVV	−28.90	7.11	a
40	TSXXPV	36.33	6.49	a
41	TXPV	−13.31	8.20	
42	TPTXS	−30.54	7.51	a
43	PTVPXVSP	−18.10	8.76	a
44	SXXVPS	−7.45	8.76	
45	PVTTTVV	8.90	8.67	
46	PTVVVV	−36.00	8.41	a
47	VSTXVVS	−27.62	7.18	a
48	TXXTVPT	22.76	8.94	a
49	PXPVXVTT	−19.08	8.34	a
50	VPXTVV	−14.33	7.86	

Note. Items 1 through 25 are grammatical and items 26 through 50 are ungrammatical. Items shown in bold served as both training and test items.
[a] Items that had a mean rating that differed reliably from zero.

where the dimensionality of $\mathbf{z}$ is equal to the dimensionalities of $\mathbf{x}$ and $\mathbf{y}$. Figure 2 depicts the operation. Circular convolution is commutative, distributes over addition, and preserves similarity.

Commutativity implies symmetric association so that the representation of a bigram *AB* would be treated as equal to the representation of a bigram *BA*. Because people encode letter strings from left to right (i.e., $AB \neq BA$), circular convolution's commutative property is undesirable. To solve the issue, we use noncommutative circular convolution. Noncommutative circular convolution is accomplished by scrambling the indices of the letter vectors before applying circular convolution to them (Jones & Mewhort, 2007; Plate, 1995). Noncommutative circular convolution is neither commutative nor associative. However, it distributes over addition and preserves similarity.

In the remainder of this article, we denote noncommutative circular convolution using an asterisk (e.g., $\mathbf{z} = \mathbf{x}^*\mathbf{y}$). For brevity, we use the term *convolution* in place of noncommutative circular convolution.

The vector returned by convolution is unique from its constituents. Thus, $\mathbf{a}^*\mathbf{b}$ and $\mathbf{a}^*\mathbf{c}$ are unique vectors (i.e., orthogonal in expectation). Because we are using noncommutative convolution, $\mathbf{a}^*\mathbf{b}$ and $\mathbf{b}^*\mathbf{a}$ are also unique. The uniqueness property extends to all orders of convolution. For example, $\mathbf{a}$ is orthogonal in expectation to $\mathbf{a}^*\mathbf{b}$, which is orthogonal in expectation to $\mathbf{a}^*\mathbf{b}^*\mathbf{c}$, and so on.

To illustrate complete string encoding with convolution, consider a string *ABCD*. First, we generate a vector for each letter, $A = \mathbf{a}$, $B = \mathbf{b}$, $C = \mathbf{c}$, and $D = \mathbf{d}$. Second, we encode the string by summing it and all of its subunits to a single vector: $\mathbf{a} + \mathbf{b} + \mathbf{c} + \mathbf{d} + \mathbf{a}^*\mathbf{b} + \mathbf{b}^*\mathbf{c} + \mathbf{c}^*\mathbf{d} + \mathbf{a}^*\mathbf{b}^*\mathbf{c} + \mathbf{b}^*\mathbf{c}^*\mathbf{d} + \mathbf{a}^*\mathbf{b}^*\mathbf{c}^*\mathbf{d}$. Of course, the example is extreme. People do not encode all information in a string, nor do they encode the same information at different encounters with it (see Wright & Whittlesea, 1998). To acknowledge these facts, we represent a letter string as a sum of g randomly sampled units, where the unit sizes range from 1 to k. Thus, in our simulations *ABCD* might be represented as $\mathbf{a} + \mathbf{b} + \mathbf{a}^*\mathbf{b} + \mathbf{b}^*\mathbf{c}^*\mathbf{d}$ at one encounter and $\mathbf{b} + \mathbf{c} + \mathbf{a}^*\mathbf{b} + \mathbf{b}^*\mathbf{c}^*\mathbf{d}$ at another, where $g = 4$ and $k = 3$. The method acknowledges encoding variability as well as the fact that we cannot identify our participants' private recoding strategies.

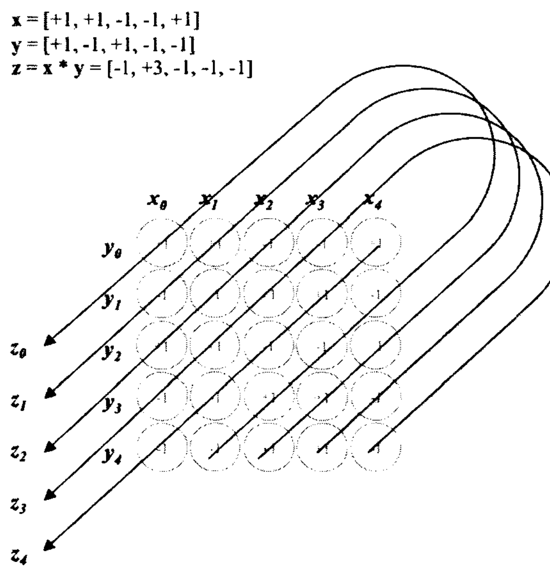

Figure 2. The figure shows two vectors, $\mathbf{x}$ and $\mathbf{y}$ (both of dimensionality $n = 5$). The outer-product of $\mathbf{x}$ and $\mathbf{y}$ is an $n \times n$ matrix. The arrows show how the elements of the outer-product matrix are summed during circular convolution to produce a summary vector $\mathbf{z}$.

Storage and Retrieval

The HEM's storage and retrieval operations follow from Hintzman's (1986) MINERVA 2 model of human memory. Thus, memory in the HEM is an m by n matrix, **M**, where m is the number of independent traces in the matrix and n is the number of features in each trace. Imperfect encoding is simulated by resetting a proportion of elements in **M** to zero (indicated data loss). The amount of loss is controlled by a parameter L that specifies the probability of storing a feature in memory correctly; thus, each element in **M** has a probability of $1 - L$ of reverting to zero.

Retrieval follows a resonance metaphor. When a probe is presented to memory, it activates all traces in parallel. Each trace's activation is a nonlinear function of its match to the probe. In the model, the activation of trace i, a_i, is computed as

$$a_i = \left(\frac{\sum_{j=1}^{n} p_j \times M_{ij}}{\sqrt{\sum_{j=1}^{n} p_j^2}\sqrt{\sum_{j=1}^{n} M_{ij}^2}} \right)^3, \tag{2}$$

where **p** is the probe, **M** is memory, i indexes the $1 \ldots m$ traces in memory, and j indexes the $1 \ldots n$ elements in the probe and memory matrix. The nonlinearity in retrieval is introduced by raising the similarity metric (the term inside the brackets on the right side of equation 2) by an odd numbered exponent. Because the exponent is odd, the sign of the similarity metric is retained in the transformation to activation.

The information that is retrieved from memory is a vector, **c**, that is called the echo. The echo is a sum of the activated traces, where each trace's contribution to the sum is in proportion to its activation by the probe. Thus, the echo is computed as

$$c_j = \sum_{i=1}^{m} a_i \times M_{ij} \; \{\text{for } j = 1 \ldots n\}, \tag{3}$$

where **c** is the echo, a_i is the activation of trace i, **M** is memory, i indexes the $1 \ldots m$ rows (i.e., traces) in memory, and j indexes the $1 \ldots n$ columns (i.e., stimulus features) in both the echo and memory matrix.

Judgment of grammaticality is predicted by echo intensity, I, that indexes the match between the probe, **p**, and the echo, **c**:

$$I = \frac{\sum_{j=1}^{n} p_j \times c_j}{\sqrt{\sum_{j=1}^{n} p_j^2}\sqrt{\sum_{j=1}^{n} c_j^2}}, \tag{4}$$

where j indexes the $1 \ldots n$ columns (i.e., stimulus features) in both the probe and the echo.

In summary, a representation of each studied trace is recorded to memory—where the representation of a stimulus is a sum of g randomly sampled subunits from sizes $1 \ldots k$. At test, a probe retrieves an aggregate of the studied traces (i.e., the echo) and judgment of grammaticality is predicted by the probe's match to the echo.

Simulation

We applied the HEM to the materials and design of Experiment 1. Each simulation included four steps. First, a random vector was generated for each unique letter in the training and test sets. Second, a representation was developed for each item in the training and test sets—different random representations were formed for the two instances of each of the six grammatical items that appeared first in the training phase and once again in the test phase. Third, the representation of each training string was stored to memory. Fourth, the echo intensity for each test item was computed and recorded. We conducted 100 independent simulations of the procedure from Experiment 1. Reported means are averaged over the 100 independent simulations.

Like our participants, the model successfully discriminated the grammaticality of test strings. The mean echo intensity for grammatical items (M = .595, SE = .002) was reliably greater than the mean echo intensity of ungrammatical items (M = .561, SE = .006). An independent samples t test confirmed that the difference was reliable, $t(48) = 5.48$, $d = 1.67$, $p < .05$. Critically, the model achieved this level of performance without grammatical knowledge.

Figure 3 shows the relationship between the mean echo intensities (simulation) and mean ratings (data) over the 50 test items. The correlation between the empirical and simulated estimates over all 50 test items was high, $r(48) = .68$, $p < .05$. The correlations for the grammatical and ungrammatical items were also high when considered independently, $r(23) = .50$ and .55, respectively, both $ps < .05$. The model's fit to participants' judgments over only those 31 items for which participants' decisions differed reliably from zero and was even higher than was the fit for the total set of 50 items, $r(48) = .78$, $p < .05$.

Participants rated the six studied grammatical test items positively. However, they did not rate those items as abnormally

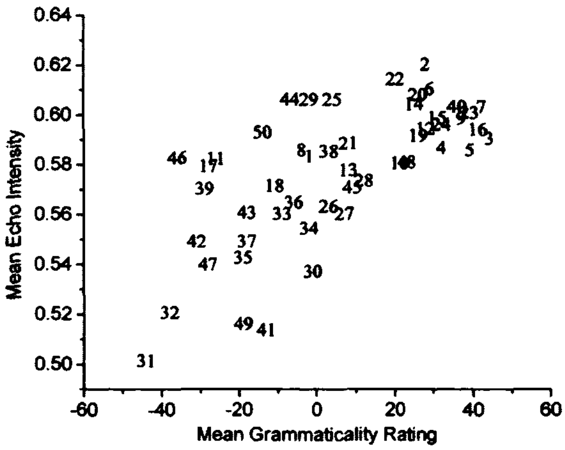

Figure 3. The model's ratings plotted against peoples' ratings of grammaticality for the 50 test items in Experiment 1. The model's judgments are expressed as mean echo intensity and peoples' ratings are expressed as mean rating of grammaticality. The parameters for the simulation were g = 3, k = 3, and L = 0.7.

positive (i.e., relative to the other unstudied grammatical test items). The model mimicked this aspect of participants' judgments: see items 2, 6, 15, 19, 20, and 23 in Figure 3. We conclude that, like our participants, the model rated the studied grammatical test items highly but did not recognize them.

Whereas participants' ratings for most test items agreed with the items' true grammatical statuses, they rated items 11, 18, 40, and 48 in disagreement with their grammatical statuses. As shown in Figure 3, the model tracked some of these details. The model rated item 18 as the worst grammatical string and rated item 11 as the third worst grammatical string. The model rated item 40 as the second best ungrammatical string but rated several of the ungrammatical items higher than item 48.

We conclude that the model does a good job of fitting peoples' judgments for individual items in an artificial grammar task. The pattern of ratings in Table 1 is consistent with an account of performance that follows from memory-based inference about grammatical status.

General Discussion

After studying grammatical training exemplars, participants can discriminate grammatical from ungrammatical test items. According to an exemplar-based account of performance, judgment of grammaticality is a retrospective inference from memory of items in the training set. Test strings that resemble the training items are classified as grammatical; test strings that differ from the training items are classified as ungrammatical.

We have described a computational instantiation of the exemplar position. The model we have proposed blends the holographic representation component from Jones and Mewhort's (2007) semantics model with the storage and retrieval components from Hintzman's (1986) MINERVA 2 model. In previous work we showed that adding the holographic representation assumptions helps the model to escape problems encountered by its previous iteration. In the present work, we showed that the holographic model does more than solve those problems: it fits performance for individual items.

Noncommutative circular convolution is an excellent tool for modelling learning in the artificial grammar task. First, it supports encoding of short-range serial order information within strings, without including knowledge of distant nonadjacent dependencies. Thus, the representation technique makes is possible to separate out the kinds of information that the model extracts from training and test exemplars. Second, because noncommutative circular convolution distributes over addition, information about short-range dependencies can be superposed into a single memory trace (i.e., the hologram). Thus, it is possible to encode items as sums of subunits. Finally, the method has a long history in psychology and so ties an explanation of performance in the artificial grammar task to explanations of memory in general including recent work in semantic representation (Cox et al., 2011; Eliasmith, 2004; Hannagan et al., 2011; Jones & Mewhort, 2007; Murdock, 1982, 1995).

Because the HEM uses different representation assumptions from the MINERVA 2 model, one might argue that we have generated a new model. That conclusion would be largely wrong. The two models are identical in terms of storage and retrieval. The only difference in the two accounts is in the information that they assume people notice in the training and test items. Jamieson and Mewhort's (2009, 2010) original adaptation of the MINERVA 2 model assumes that people notice the spatial relations between all letters in a string. The holographic model, on the other hand, assumes that people notice the short-range serial order information in strings (i.e., single letters, bigrams, and trigrams). Whereas both representations are valid—people can encode any of several kinds of information in training and test strings—only the serial-order representation in the convolution model agrees with standard empirical facts. Participants favor regularities in adjacent elements over regularities in nonadjacent elements (Kinder, 2010), and only in special circumstances do participants learn nonadjacent dependencies (see Johnstone & Shanks, 2001). Because the convolution-based model agrees with empirical facts, we favor it.

Jamieson and Mewhort (2010, Table 8) presented item level data from another artificial grammar experiment like the one presented here. We fit the convolution model to those point estimates and once again obtained strong fits. First, the mean echo intensity for grammatical strings was greater than the mean echo intensity for ungrammatical strings, thus confirming that the model distinguishes grammatical from ungrammatical strings without grammatical knowledge. More critically, the correlation between the empirical and simulated estimates over the full set of 50 test items was high, $r(48) = .67$, $p < .05$, as were the correlations for the grammatical and ungrammatical items considered independently, $r(23) = .52$ and $.51$, respectively, both $ps < .05$. Although the strong fits to those data do nothing more than reinforce our conclusions, it is important to note that the model fits item level data in more than one experiment. To our knowledge, no others have published item level data. So, the data presented here, and the data presented in Jamieson and Mewhort (2010), present the best databases for model evaluation of item level predictions.

Our account denies that participants learn rules and regularities in the training list. Nevertheless, the model discriminates grammatical from ungrammatical strings. What, then, enables the model to behave as if it knows the grammar when it does not?

The HEM assumes that during retrieval, information in memory is collapsed to an echo. Even with undifferentiated activation of traces (i.e., where $a_i = 1$ for all $i = 1 \ldots m$ traces in memory), the echo highlights commonalities in the studied exemplars. If traces share a particular feature, then that feature will figure prominently in the echo. However, because traces in memory are distributed data structures, the echo will highlight many features. Thus, whereas memory contains no information about rules and regularities, the echo suffices.

To appreciate the behavior and flexibility of the echo, consider the influence of differential activation in memory by a test probe. Whereas memory might hold four items—*MTRRV*, *MMRVV*, *TPRXT*, and *TPXRT*—it enables a great many more summaries of structure in the items. For example, when presented with *MTTRV* the echo will reflect traces one and two most strongly. When presented with *TXRTP*, the echo will reflect traces three and four most strongly. When presented with *MTRXT*, the echo will reflect all four traces equally and generate a representation intermediate to the echoes for *MTTRV* and *TXRTP*. In short, memory of training exemplars carries no particular summary of structure in the training list but, rather, carries several potential responses specific to how memory is queried. Understanding that memory does not extract and store a particular abstraction over experience, but rather, holds items with the *potential* to do so at retrieval is key to understanding how the HEM accounts for performance in the artificial grammar task (see Vokey & Higham, 1999). Of course, the position that knowledge emerges in retrieval even though that knowledge is not represented directly in memory also explains how people manage to judge the grammaticality of test items well, even without an ability to express the grammar.

Memory-based accounts of performance, like the one presented here, contrast with learning-based accounts. In a learning account, participants prepare for the test by learning rules and regularities. At test, performance is a reflection of success at learning (e.g., Knowlton & Squire, 1996; Reber, 1967). In a memory-based account, participants store the training exemplars. At test, performance reflects how the participants use memory of training items to infer grammaticality. Both types of account explain judgment of grammaticality. However, a learning account is forced to answer questions that the memory account is not. Why do participants learn rules and regularities that allow them to judge the grammaticality of test probes? How do they know which rules and regularities to learn? If they know the rules and regularities, why can't they express that knowledge?

If one assumes that participants set out to learn rules and regularities, much of the problem would appear to fade away (although, Wright & Whittlesea, 1998, would ask how the subject knows which rules and regularities are relevant). However, even if this were the case, it would not benefit performance. In cases where the experimenter tells participants to study training items in preparation for a test of grammaticality, performance gets worse rather than better (e.g., Reber, Kassin, Lewis, & Cantor, 1980). Moreover, the assumption that participants learn the rules and regularities contradicts the nature of the research problem. A critical and defining feature of an implicit learning task is that the experimenter withholds and even disguises the test—at least until it is too late for the participant to engage in active and deliberate learning. Thus, claiming that participants learn the structure of the materials in anticipation of performing judgment of grammaticality necessitates a claim for compulsive and obligatory learning of rules and regularities in a training list. The memory approach escapes the problem. Subjects learn the exemplars. At test, they infer grammaticality by a test item's similarity to the studied set. The ability to discriminate grammatical status is an incidental benefit.

Simon (1969) argued that sophisticated behavior could emerge from unsophisticated systems when those systems are made to operate in a complex environment (Morgan, 1903; Todd & Gigerenzer, 2007). Accordingly, he warned psychologists to resist attributing complex behaviour to complex mechanisms. Tero et al. (2010) recently published a curious illustration of Simon's position. A slime mold (*Physarum polycephalum*) placed in a structured environment formed transport networks that rivaled human engineered transport networks by measures of efficiency, fault tolerance, and cost. In discussing the result, one might call the slime mold sophisticated. However, analysis reveals that its behavior follows from the application of a simple foraging algorithm to a complex environment (see Tero et al.'s model for a description of the algorithm). We see peoples' performance in the artificial grammar as another illustration of Simon's point. In discussing peoples' ability to discriminate grammatical from ungrammatical items, one might be tempted to call the participant sophisticated. However, analysis shows that the participant's decisions are explicable by the operation of a basic mechanism operating in a complex environment (see our model for a description of the algorithm). Contrary to appearances, peoples' solution to the artificial grammar task is more elegant than it is sophisticated.

Résumé

La recherche consistait à appliquer un modèle de mémoire à multiples traces pour expliquer la performance durant une tâche de grammaire artificielle. Le modèle combine la méthode de convolution pour la représentation, du modèle BEAGLE, de Jones et Mewhort de la mémoire sémantique (Jones, M. N., & Mewhort, D. J. K. [2007]. Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114,* 1–37) et le modèle d'encodage à multiples traces et de récupération MINERVA 2, de D. L. Hintzman (1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428) de la mémoire épisodique. L'article décrit une expérience de grammaire artificielle dans le cadre de laquelle le modèle est appliqué aux données pour chacun des items. Il est avancé que la performance de la tâche de grammaire artificielle s'explique le mieux en tant que processus d'inférences rétrospectives à partir de la mémoire.

Mots-clés: modèle de référence, représentation réduite holographique, grammaire artificielle, MINERVA 2.

References

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization,* (pp. 169–211). Hillsdale, NJ: Erlbaum Inc.

Cox, G. E., Kachergis, G., Recchia, G., & Jones, M. N. (2011). Toward a scalable holographic word-form representation. *Behavior Research Methods, 43,* 602–615. doi:10.3758/s13428-011-0125-5

Dienes, Z. (1992). Connectionist and memory-array models of artificial grammar learning. *Cognitive Science, 16,* 41–79. doi:10.1207/s15516709cog1601_2

Dienes, Z., Broadbent, D. E., & Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 875–887. doi:10.1037/0278-7393.17.5.875

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgement: How conscious and how abstract? *Journal of Experimental Psychology: General, 113,* 541–555. doi:10.1037/0096-3445.113.4.541

Eliasmith, C. (2004). Learning context sensitive logical inference in a neurobiological simulation. In S. Levy & R. Gayler (Eds.), *Compositional connectionism in cognitive science* (pp. 17–20). Menlo Park, CA: AAAI Press.

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

Gabor, D. (1969). Associative holographic memories. *IBM Journal of Research and Development, 13,* 156–159.

Hannagan, T., Dupoux, E., & Christophe, A. (2011). Holographic string encoding. *Cognitive Science, 35,* 79–118. doi:10.1111/j.1551-6709.2010.01149.x

Higham, P. A. (1997). Chunks are not enough: The insufficiency of feature frequency-based explanations of artificial grammar learning. *Canadian Journal of Experimental Psychology, 51,* 126–138. doi:10.1037/1196-1961.51.2.126

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428. doi:10.1037/0033-295X.93.4.411

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, 36,* 1529–1535. doi:10.1037/a0020598

Jamieson, R. K., & Mewhort, D. J. K. (2005). The influence of grammatical, local, and organizational redundancy on implicit learning: An analysis using information theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 9–23. doi:10.1037/0278-7393.31.1.9

Jamieson, R. K., & Mewhort, D. J. K. (2009). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology, 62,* 550–575. doi:10.1080/17470210802055749

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. doi:10.1080/17470210903267417

Jamieson, R. K., & Mewhort, D. J. K. (2011). Grammaticality is inferred from global similarity: A reply to Kinder (2010). *Quarterly Journal of Experimental Psychology, 64,* 209–216. doi:10.1080/17470218.2010.537932

Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology, 42,* 61–112. doi:10.1006/cogp.2000.0743

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114,* 1–37. doi:10.1037/0033-295X.114.1.1

Kinder, A. (2010). Is grammaticality inferred from global similarity? Comment on Jamieson & Mewhort (2009). *Quarterly Journal of Experimental Psychology, 63,* 1049–1056. doi:10.1080/17470211003718713

Kinder, A., & Shanks, D. R. (2001). Amnesia and the declarative/nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience, 13,* 648–669. doi:10.1162/089892901750363217

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. doi:10.1037/0278-7393.22.1.169

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

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). The role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1083–1100. doi:10.1037/0278-7393.15.6.1083

Metcalfe-Eich, J. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology, 56,* 485–491.

Morgan, C. L. (1903). *An introduction to comparative psychology.* London, UK: Walter Scott Publishing.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626. doi:10.1037/0033-295X.89.6.609

Murdock, B. B. (1995). Developing TODAM: Three models for serial-order information. *Memory & Cognition, 23,* 631–645. doi:10.3758/BF03197264

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9,* 247–255. doi:10.1111/1467-9280.00051

Perruchet, P., & Gallego, J. (1997). A subjective unit formation account of implicit learning. In D. C. Berry (Ed.), *How implicit is implicit learning? Debates in psychology* (pp. 124–161). Oxford, UK: Oxford University Press.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119,* 264–275. doi:10.1037/0096-3445.119.3.264

Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks, 6,* 623–641. doi:10.1109/72.377968

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

Pothos, E. M., & Bailey, T. M. (2000). The importance of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 847–862. doi:10.1037/0278-7393.26.4.847

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior, 6,* 855–863. doi:10.1016/S0022-5371(67)80149-X

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

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit models in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 492–502. doi:10.1037/0278-7393.6.5.492

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 592–608. doi:10.1037/0278-7393.16.4.592

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Tero, A., Takagi, S., Saigusa, T., Ito, K., Bebber, D. P., Fricker, M. D., . . . Nakagaki, T. (2010). Rules for biologically inspired adaptive network design. *Science, 327,* 439–442. doi:10.1126/science.1177894

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science, 16,* 167–171. doi:10.1111/j.1467-8721.2007.00497.x

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 328–344. doi:10.1037/0278-7393.18.2.328

Vokey, J. R., & Higham, P. A. (1999). Implicit knowledge as automatic, latent knowledge. *Behavioral and Brain Sciences, 22,* 787–788. doi:10.1017/S0140525X99582186

Wright, R. L., & Whittlesea, B. W. A. (1998). Implicit learning of complex structures: Active adaptation and selective processing in acquisition and application. *Memory & Cognition, 26,* 402–420. doi:10.3758/BF03201149

Received August 12, 2011
Accepted December 21, 2011 ■