Recoding and representation in artificial grammar learning

Chrissy M. Chubala · Randall K. Jamieson

Abstract We apply an exemplar model of memory to explain performance in the artificial grammar task. The model blends the convolution-based method for representation developed in Jones and Mewhort's BEAGLE model of semantic memory (Psychological Review 114:1–37, 2007) with the storage and retrieval assumptions in Hintzman's MINERVA 2 model of episodic memory (Behavior Research Methods, Instruments, and Computers, 16:96–101, 1984). The model captures differences in encoding to fit data from two experiments that document the influence of encoding on implicit learning. We provide code so that researchers can adapt the model and techniques to their own experiments.

Keywords Artificial grammar learning · Mental organization · Exemplar model · Holographic reduced representation

In an artificial grammar task, participants study letter strings constructed according to the rules of an artificial grammar. Afterward, they attempt to discriminate novel grammatical from novel ungrammatical test strings. Typically, people can discriminate the two classes of test items, but they cannot articulate the grammar (Reber, 1967).

Three kinds of theories have been proposed to explain peoples' performance in the artificial grammar task. *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 designated implicit (e.g., Mathews et al., 1989; Reber, 1967). *Statistical theories* propose that participants learn regularities in the training list (e.g., letter frequencies)

and, at test, endorse those strings that exhibit the learned regularities. The statistical theories split 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 training exemplars in memory and, at test, endorse strings that remind them of the studied exemplars. Because no implicit knowledge of the grammar is assumed, the discrepancy between participants' performance and awareness is irrelevant (e.g., Brooks, 1978; Higham, 1997; Jamieson, Holmes, & Mewhort, 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 by extending Hintzman's (1984, 1986) MINERVA 2 model of human memory. According to the account, participants store each exemplar as a separate trace in memory. When a test probe is presented, each trace is activated in proportion to its similarity to the probe. Judgment of grammaticality is based on the similarity of the probe to a weighted sum of the activated traces. In short, a probe that retrieves a good representation of itself from memory is judged grammatical; a probe that does not retrieve a good representation of itself from memory is judged ungrammatical. The account has roots in Brooks's (1978; Vokey & Brooks, 1992) instance-based analysis of implicit learning.

Following Brooks (1978), Jamieson and Mewhort's (2009, 2010) model encoded exemplars as whole strings. Specifically, each letter was represented by a unique vector of random elements, and each letter string was represented by concatenating letter vectors in corresponding order to the stimulus. For example, a string *ABCD* was represented by first generating a vector for each letter—**a, b, c,** and **d**—and then, concatenating the letter vectors **a**//**b**//**c**//**d**, where // indicates concatenation.

C. M. Chubala (✉) · R. K. Jamieson
Department of Psychology, University of Manitoba,
Winnipeg, MB, Canada R3T 2N2
e-mail: umchubal@cc.umanitoba.ca

Although the concatenation-based model simulated performance in several artificial grammar experiments, Kinder (2010) soon pointed out that whole-string encoding was generally at odds with the chunk-based encoding that people tend to engage in when studying training strings in the artificial grammar task (Perruchet & Pacteau, 1990). For example, people more often encode a string *RGBRGBPY* as "*RGB, RGB, PY*" than as "*RGBRGBPY*" (Jamieson & Mewhort, 2005).

To resolve the issue, Jamieson and Mewhort (2011) reengineered the representation assumptions in their model on the basis of a scheme for holographic representation developed in Jones and Mewhort's (2007; Plate, 1995) BEAGLE model of semantic memory. In the new model, letters were represented as vectors of random elements, letter subsequences (i.e., bigrams, trigrams) were constructed by applying noncommutative circular convolution to the letter vectors, and letter strings were represented by summing encoded subsequences into a single vector. Thus, whereas the concatenation-based model assumed that a string is remembered as a whole item with letters tied to serial positions, the holographic model assumed that a string is remembered as a sum of encoded subunits (i.e., single letters, bigrams, trigrams, etc.).

Jamieson and Mewhort (2011; Jamieson & Hauri, 2012) showed that the convolution-based model fit data that its concatenation-based predecessor did not (Kinder, 2010). On the basis of that success, they argued that holographic representation is a better method to represent the information that people store about training and test strings and, therefore, provides a more accurate account of learning in the artificial grammar task. There was, however, a limitation in how the model was applied.

In most artificial grammar experiments, participants are granted freedom to encode training and test strings as they choose. Thus, to accommodate the fact that they could not know a participant's private encoding strategies, Jamieson and Mewhort (2010) adopted an assumption of random, rather than structured, sampling of subunits from training and test strings. For example, when presented with a string *RBGRGBPY*, it was assumed that participants and, thus, the model employed a random sampling of units. Whereas the adoption of random sampling was a necessary step to model peoples' performance with free encoding, Jamieson and Mewhort (2010) speculated that their holographic encoding scheme could be used to predict the influence of structured encoding. This article follows up on this claim.

In the work that follows, we apply the holographic exemplar model (HEM) to data from an artificial grammar experiment in which peoples' encoding was brought under experimental control. However, we begin by applying the model to an implicit rule-learning procedure reported by Wright and Whittlesea (1998).

Wright and Whittlesea (1998)

Wright and Whittlesea (1998) reported an experiment that showed a predictable influence of encoding on performance in a hidden rule-learning task. In the experiment, participants read digit strings, all of which were four digits in length and all of which conformed to an *odd–even–odd–even* rule. For example, 1234 was a permissible string, but 1235 was not. Participants were assigned to one of two encoding conditions. Participants assigned to a *ones* condition read each training string aloud as successive digits (e.g., 1234 was read as "one, two, three, four"). Participants assigned to a *tens* condition read each training string aloud as successive pairs (e.g., 1234 was read aloud as "twelve, thirty-four"). Following training, participants in both encoding conditions performed two-alternative forced choice recognition for the studied strings; unknown to participants, all test strings were new, with one string in each test pair consistent with the odd–even–odd–even rule and the other inconsistent with the odd–even–odd–even rule. Wright and Whittlesea argued that if participants implicitly learn and use the odd–even rule (i.e., automatic rule learning), performance in the two groups ought to be identical. However, if encoding influences what is learned, performance should differ in the two conditions.

The top panel in Fig. 1 shows the results of the experiment. As is shown, participants in the tens condition discriminated rule-bound from rule-violating test strings better than did participants in the ones condition, $p < .05$. The difference is consistent with the position that encoding influences performance and contradicts the position that performance is under the control of implicit learning. For our purposes, Wright and Whittlesea's (1998) data illustrate that encoding influences judgments at test and offers a concrete target to evaluate the holographic model's ability to accommodate differences in performance as a function of differential encoding.

We now describe Jamieson and Mewhort's (2011) HEM. Afterward, we apply the model to Wright and Whittlesea's (1998) experiment. If the model offers a capable account of differential encoding in the ones and tens conditions, it will predict better discrimination of rule-conforming from rule-violating items in the tens than in the ones condition.

The holographic exemplar model

Jamieson and Mewhort's (2011) HEM merges the holographic representation scheme from Jones and Mewhort's
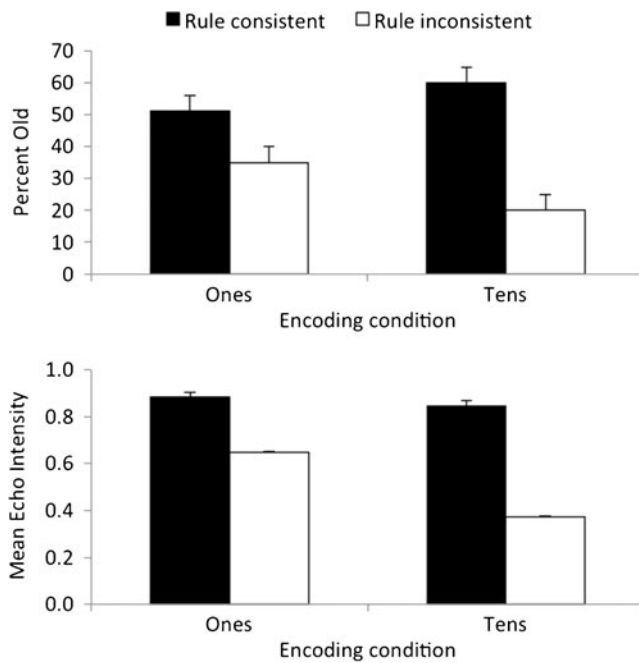
Fig. 1 The top panel shows participants' performance in Wright and Whittlesea's (1998) Experiment 3B. The bottom panel shows the holographic exemplar model's fit to the results. Whiskers on both graphs are standard errors

$\mathbf{x} = [+1, +1, -1, -1, +1]$
$\mathbf{y} = [+1, -1, +1, -1, -1]$
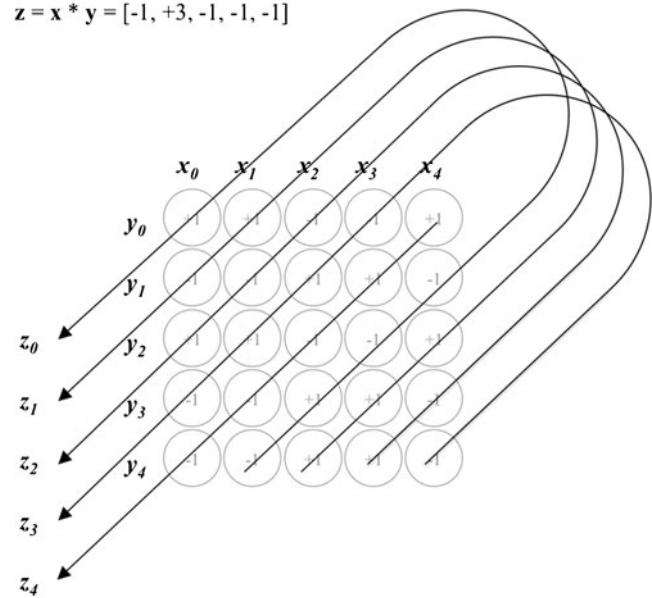$\mathbf{z} = \mathbf{x} * \mathbf{y} = [-1, +3, -1, -1, -1]$

Fig. 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, $\mathbf{z}$

(2007) BEAGLE model of semantic memory with the storage and retrieval model from Hintzman's (1984, 1986) MINERVA 2 model of episodic memory.

Representation

In the HEM, a letter is an n dimensional vector. Each element in a letter vector takes a random value sampled from a normal distribution with mean zero and variance $1/\sqrt{n}$. Letter subsequences are formed by applying noncommutative circular convolution to letter vectors. Letter strings are formed by summing encoded letter subsequences.

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

$$z_i = \sum_{j=0}^{n-1} x_j \times y_{(i-j) \bmod n} \quad \{for\ i = 0\ to\ n-1\} \tag{1}$$

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

Commutativity implies symmetric association. Thus, the representation of a bigram AB is equal to the representation of a bigram BA. Because people tend to encode symbols from left to right (i.e., $AB \neq BA$), circular convolution's commutative property is undesirable. We solve the problem by using a noncommutative

version of circular convolution, accomplished by scrambling the indices of the letter vectors before applying circular convolution to them (see Jones & Mewhort, 2007, Appendix A). Whereas noncommutative circular convolution distributes over addition and preserves similarity, it is neither commutative nor associative.

From here forward, 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*.

A vector returned by convolution is orthogonal in expectation to its constituents. Thus, $\mathbf{a}$ is orthogonal in expectation to $\mathbf{a}*\mathbf{b}$, which is orthogonal in expectation to $\mathbf{a}*\mathbf{b}*\mathbf{c}$, and so on. Because we are using noncommutative convolution, the convolution of two vectors also differs depending on the order in which the vectors are convolved: $\mathbf{a}*\mathbf{b}$ is orthogonal in expectation to $\mathbf{b}*\mathbf{a}$, which is orthogonal in expectation to $\mathbf{b}*\mathbf{a}*\mathbf{c}$, and so on.

In addition to serial order information within units, convolution can be used to encode information about the order of encoded units (Cox, Kachergis, Recchia, & Jones, 2011; Hannagan, Dupoux, & Christophe, 2011). For example, using vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ to represent letters A through D and vectors $\mathbf{1}$ and $\mathbf{2}$ to represent the order of units in a string, the vector $\mathbf{a}*\mathbf{b}*\mathbf{1} + \mathbf{c}*\mathbf{d}*\mathbf{2}$ represents "*AB followed by CD*," whereas the vector $\mathbf{c}*\mathbf{d}*\mathbf{1} + \mathbf{a}*\mathbf{b}*\mathbf{2}$ represents "*CD followed by AB*." In the simulations that follow, we will use order codes to

capture the fact that participants know the order of the units they encode.

Storage and retrieval

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, where the degree 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. Prior work with the concatenation model showed that decreasing *L* impairs memory performance overall but has a greater impact on performance in tasks that require knowledge of the specific (e.g., recognition), as compared with tasks that require knowledge of the general (e.g., classification; see Jamieson et al., 2010).

Retrieval in the HEM follows a resonance metaphor. Presenting a probe vector to memory causes all traces in memory to activate 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

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

where **p** is the probe (i.e., a row vector), **M** is memory, *i* indexes the 1…*m* traces in memory, and *j* indexes the 1…*n* columns in the probe and memory matrix. Nonlinearity is introduced in retrieval by raising the similarity metric (the term inside brackets in Equation 2) to an odd numbered exponent. The exponential transformation ensures that retrieval is selective to the traces that match the probe most closely.

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

$$c_j = \sum\limits_{i=1}^{m} a_i \times M_{ij} \quad \{for\ j = 1 \dots n\} \qquad (3)$$

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

Judgment of grammaticality is predicted by a scalar, *I*, called *echo intensity*, that indexes the match between the probe, **p**, and the echo, **c**. Echo intensity is computed as

$$I = \frac{\sum\limits_{j=1}^{n} p_j \times c_j}{\sqrt{\sum\limits_{j=1}^{n} p_j^2} \sqrt{\sum\limits_{j=1}^{n} c_j^2}}, \qquad (4)$$

where *j* indexes the 1…*n* columns (i.e., stimulus features) in both the probe and the echo. In the simulations that follow, results are presented as mean echo intensities across multiple simulated participants. An item is endorsed to the extent that it can be reproduced from memory.

Simulation of Wright and Whittlesea (1998)

We conducted 100 independent simulations of the ones encoding condition and 100 independent simulations of the tens encoding condition. Each independent simulation followed the same procedure. First, we generated a unique random vector for each digit 1 through 8. Second, we encoded each of the training and test strings depending on the relevant encoding scheme. In the ones encoding condition, strings were encoded by summing the convolution of each digit with its position. In the tens encoding condition, strings were encoded by convolving the first two digits in a string along with order vector 1 and the second two digits in a string along with order vector 2. For example, the string 1234 (i.e., rewritten as *ABCD* here for clarity of notation) was encoded as **a*1 + b*2 + c*3 + d*4** in the ones condition and as **a*b*1 + c*d*2** in the tens condition. Third, the training strings were stored to memory, *L* = 1.0. Fourth, we computed the echo intensity for each of the test items. Because Wright and Whittlesea (1998) did not provide their stimulus materials, we generated a set based on their described algorithm (see Table 1). Code for the simulation can be downloaded from our Web site.[1]

Simulation results are presented in the bottom panel of Fig. 1. Like Wright and Whittlesea's (1998) participants, the HEM discriminated rule-consistent from rule-inconsistent test items in both the ones and tens encoding conditions but discriminated better in the tens than in the ones condition. However, mean echo intensity for grammatical items in the tens condition is slightly lower, rather than slightly higher, as compared with mean echo intensity for grammatical items in the ones condition. A comparison of the upper and lower panels shows that this is at odds with the experimental data.

We conducted additional simulations with random, rather than structured, sampling of subunits (i.e., the method used in

Table 1 Materials used in the simulation of Wright and Whittlesea's (1998) experiment

Training exemplars	Test exemplars	
	Rule bound	Rule violating
1216	1612	*2*418
1234	3412	1*5*32
1638	3816	18*4*6
1652	5216	185*7*
3458	5834	*4*256
3474	7434	3*5*72
3876	7638	36*4*8
3812	1238	361*5*
5216	1652	*6*418
5234	3452	5*7*32
5838	3858	56*4*6
5852	5258	565*7*
7458	5874	*8*256
7476	7674	7*3*78
7674	7476	78*6*2
7612	1276	781*5*

Italicized/bolded digits are rule violations

previous applications of the HEM). Those simulations failed to produce Wright and Whittlesea's (1998) data; discrimination of rule-consistent from rule-inconsistent test strings was equivalent in the ones and tens conditions. We conclude that the HEM can be used to capture differences in peoples' performance as a function of differential structured encoding.

Although the HEM captures Wright and Whittlesea's (1998) data to show that encoding influences performance in an implicit rule-learning task, Wright and Whittlesea's rule-learning procedure differs in several important ways from an artificial grammar task. First, whereas Wright and Whittlesea tested participants' learning of a simple rule, based on the odd/even concept, participants in a standard artificial grammar task are tested on a complex network of rules defined by a sequential contingency structure. Second, whereas Wright and Whittlesea did not inform their participants that the training materials were constructed using a rule and asked their participants to perform recognition, participants in a standard artificial grammar task are informed that training items were constructed according to rules and are asked to classify novel test strings as either consistent or inconsistent with those rules. Finally, whereas Wright and Whittlesea tested performance with a two-alternative forced choice test, participants in a standard artificial grammar task judge the grammatical status of each test string independently. For all three reasons, our analysis of Wright and Whittlesea's experiment does not speak directly to the problem of artificial grammar learning. To address the issue, we report an artificial grammar experiment to match the experimental structure in Wright and Whittlesea's implicit rule-learning study.

Experiment

We designed an artificial grammar task to match key features from Wright and Whittlesea's (1998) implicit rule-learning task. In a training phase, participants studied 30 grammatical training strings in one of two encoding conditions. At test, they rated the grammaticality of 40 unstudied test strings: 20 grammatical and 20 ungrammatical. Participants assigned to a bigrams-encoding condition were shown training strings in three successive stages (e.g., *MQRVTX* was presented as *MQ* for 1,666 ms, *MQRT* for 1,666 ms, and *MQRVTX* for 1,668 ms). Participants assigned to a trigrams-encoding condition were shown training strings in two successive stages (e.g., *MQRVTX* was presented as *MQR* for 2,500 ms and *MQRTVX* for 2,500 ms). Figure 3 shows the presentation methods.[2] Despite the differences between our artificial grammar experiment and Wright and Whittlesea's rule-learning experiment, both ask the same question: Does manipulating how participants encode materials in training impact their judgments of structural consistency at test?

Method

Participants

Ninety-two undergraduate students from the University of Manitoba participated in the experiment. Half of the participants were assigned to the bigram-encoding condition, and half to the trigram-encoding condition. All participants reported normal or corrected-to-normal vision.

Apparatus

The experiment was administered on personal computers (PCs). Each PC was equipped with a 22-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 using the keyboard to report the rules of the grammar.

Materials

Stimuli were strings of six letters. Only the letters *M, Q, R, T, V,* and *X* were used. Sequential rules were used to

---

[2] The timing used in the experiment equates the overall presentation time of each training string in the bigrams and trigrams conditions but does not equate the presentation time per unit (i.e., bigram and trigram) in the two conditions. Unfortunately, there is no way to simultaneously equate both factors.
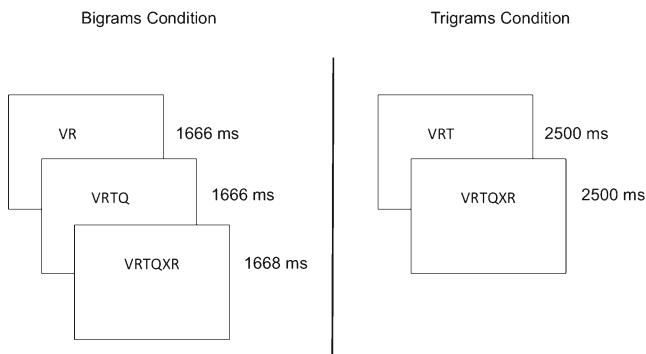
Bigrams Condition   Trigrams Condition

VR	1666 ms
VRTQ	1666 ms
VRTQXR	1668 ms

| VRT | 2500 ms |
| VRTQXR | 2500 ms |

Fig. 3 Description of how stimulus presentation was timed during the training phase of the bigram (left panel) and trigram (right panel) encoding conditions in the reported experiment

constrain which letters could and could not follow one another from left to right in a grammatical string.

The rule set—probabilities with which the letters M, Q, R, T, V, and X could follow one another in successive serial positions of a string—is shown in Table 2. For example, when the letter M appeared at any position in a string, the letters V and X could follow, each with probability .5, but letters M, Q, R, and T could not. Transition grammars, like the one we use here, solve problems associated with finite state grammars used in most studies of artificial grammar learning. In particular, they do not impose positional constraints; they describe the transition probabilities among symbols at all positions in a sequence; they can be used to construct a large number of strings; and because they describe only first-order symbol transitions, the number of rule violations in ungrammatical strings can be controlled and quantified (Jamieson & Mewhort, 2005).

The stimulus set for the experiment included 30 grammatical training strings, 20 grammatical test strings, and 20 ungrammatical test strings. Grammatical strings were constructed by selecting one of four letters, M, Q, V, or X, to the

Table 2 The grammar shows transition probabilities among letters M, Q, R, T, V, and X from position n of a sequence to position n + 1 of a sequence

Letter (n + 1)						
Letter (n)	M	Q	R	T	V	X
M	0.00	0.00	0.00	0.00	0.50	0.50
Q	0.50	0.00	0.00	0.00	0.00	0.50
R	0.00	0.50	0.00	0.50	0.00	0.00
T	0.00	1.00	0.00	0.00	0.00	0.00
V	0.00	0.50	0.50	0.00	0.00	0.00
X	0.00	0.00	0.33	0.33	0.33	0.00

Strings could begin with letter M, Q, V, or X but could not begin with letter R or T. Any letter could end a string. The grammar is very predictable, with a grammatical redundancy of .64 (Jamieson & Mewhort, 2005).

first position in a string and assigning a letter to each successive serial position of the string in accordance with the rules in the grammar. Ungrammatical strings were constructed by assigning one of the four letters M, Q, V, or X to the first position in a string and then assigning any of the six available letters to the remaining serial positions in the string. Rule violations in the ungrammatical strings occurred between the second and third and between the fourth and fifth letters. Table 3 gives a complete list of the training and test strings.

Procedure

Participants were tested in groups of 3–7. After participants had been seated at different computer terminals, they were told that they would be shown strings of letters and that it

Table 3 Materials used in the experiment

Training exemplars	Test exemplars	
	Grammatical	Ungrammatical
VRTQXR	XVRTQX	XTMVMX
XVQMVR	XTQXRT	QMRTMX
QMXVRQ	XVQMXV	VRVQRT
MXRQXT	XRTQXT	MXQXMV
XVQXRT	QXTQXV	XVMVTQ
MXRQXR	VRQXTQ	QXXRMV
MXVQMV	XRTQXR	VQRQVR
VQMXTQ	XTQMVR	MVTQVR
MVRQXV	QMVQXR	XTRQVQ
VQXRTQ	MVQXVR	QMMVTQ
XRTQMX	VQMVRT	VRMXQM
QXRQMX	XRTQXV	MXQMRQ
XTQMXT	VQXTQX	XRMVTQ
MXVRTQ	MXRTQX	QXMVVR
XRQMXT	MXRTQM	VQTQRT
QXRQMV	VRTQMX	MVXTVQ
QXTQMX	QMXTQX	XVTQRQ
MVRTQX	MVRQXT	XRMXQM
QXVQXR	VRTQXV	XRVQTQ
MVQXRT	QXVQXT	MXQMRT
XVRTQM		
VRTQMV		
QMXTQM		
QMXRTQ		
XRQXVR		
QXVRTQ		
MVQMXR		
MXRQMV		
VRQXRQ		
QMVRQM		

would be their job to read each one. The participants were not yet told about the judgment-of-grammaticality test that would follow.

The participant initiated the training phase by clicking on the message "Start" that was displayed at the center of the computer screen, after which the screen was cleared for 1,000 ms. Immediately thereafter, the first training string was displayed in 30-point Arial font. At a distance of 60 cm (i.e., approximately an arm's length from the computer screen), each letter had a visual angle of approximately 1.01° vertical and 0.82° horizontal.

Participants assigned to the bigram-encoding condition saw each training string presented as three successive bigrams. Participants assigned to the trigram group saw each string presented as two successive trigrams. Figure 3 gives a more detailed description of how each string was presented in the two encoding conditions.

After a string had been presented, the screen was cleared for 1,000 ms; immediately thereafter, the next string was presented. This cycle was repeated until all 30 of the training strings had been presented.

Following the training phase, the screen was cleared, and instructions for the test phase were presented. A button marked "OK" was presented beneath the instructions. The instructions informed the participant that all of the studied strings were constructed according to the rules of an artificial grammar and that he or she would now rate the grammaticality of 40 novel unstudied test strings, half of which would conform to the rules and half of which would not. The instructions told participants to raise their hand if they had any questions and, otherwise, to continue on to the test phase by clicking on the "OK" button.

On each test trial, a string was presented at the center of the screen. In contrast to the training phase, each test string was presented as a whole without any manipulation of presentation. A line approximately 5 cm in length was displayed 3 cm below the string with 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. Although participants were not shown numbers on the slider, the left extreme of the line 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.

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. This cycle continued until all of the 40 test strings had been presented.

After all of the test strings had been presented and the participant had provided a response to each one, the screen was cleared; 1,000 ms later, a text editor appeared on the computer screen. A message above the text editor invited the participant to use the computer keyboard to describe the rules that he or she thought had been used to construct the training strings. When they were finished, they clicked on a button marked "OK." When the participant clicked on the button, a message appeared thanking the participant for his or her participation in the study.

Results and discussion

Participants' use of the continuous rating scale was extremely varied. Whereas some participants used the full range of the scale, other used an extremely limited range of the scale. Because of those differences, ratings from overconfident participants overshadowed ratings from conservative participants. To solve the problem, we rescored the continuous ratings as binary decisions. Ratings greater than zero were scored as grammatical responses. Ratings less than zero were scored as ungrammatical responses.

The top panel in Fig. 4 shows the mean percentage of grammatical responses that participants gave to the grammatical and ungrammatical test strings. Performance is shown separately for the bigram- and trigram-encoding groups. As is shown, participants in both groups endorsed grammatical strings more than ungrammatical strings. However, participants in the trigram-encoding condition discriminated grammatical from ungrammatical strings better than did participants in the bigram-encoding condition. A test for the interaction between stimulus type and encoding condition confirmed the difference, $F(1, 90) = 5.27$, $MSE = 181.75$, $p < .05$.

We now apply the HEM to the materials and procedure of our experiment. If the model's fit to Wright and Whittlesea's (1998) rule-learning experiment is general, the model should track the influence of encoding in our artificial grammar experiment.

Simulation of experiment

We conducted 100 independent simulations of the bigram-encoding condition and 100 independent simulations of the trigram-encoding condition from our experiment. Each independent simulation followed the same procedure. First, we generated a unique random vector for each letter in the training and test lists. Second, we encoded each of the training and test strings using the encoding scheme that was relevant to the simulated condition. For example, the
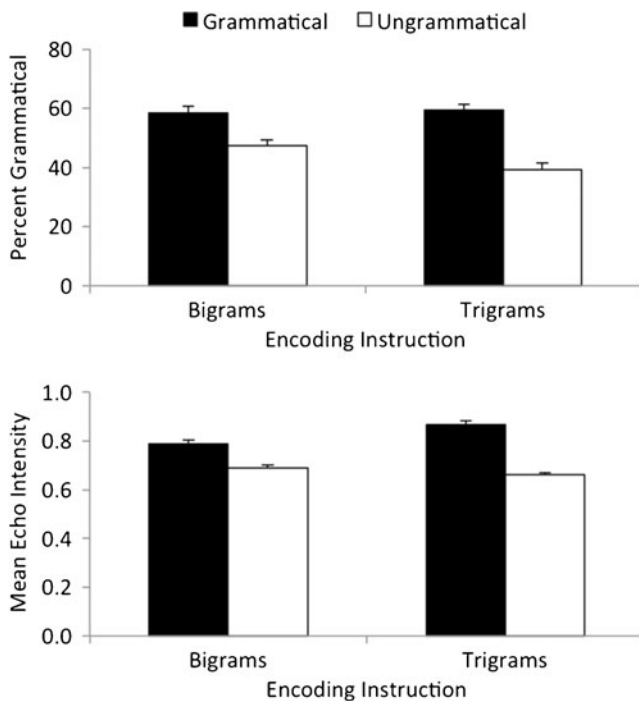
Fig. 4 The top panel shows the mean percentages of grammatical and ungrammatical test strings that participants endorsed as grammatical. Performance is shown separately for the bigram- and trigram-encoding groups. The bottom panel shows the holographic exemplar model's fit to the data. Whiskers on both graphs are standard errors

string *VRTQXR* was encoded as **v*r*1 + t*q*2 + x*r*3** (i.e., "*VR* followed by *TQ* followed by *XR*") in the bigram condition and **v*r*t*1 + q*x*r*2** (i.e., "*VRT* followed by *QXR*") in the trigram condition. Third, the training strings were stored to memory, $L = 1.0$. Fourth, we computed the echo intensity for each of the test items. Code for the simulation can be downloaded from our Web site.[3]

The lower panel in Fig. 4 shows mean echo intensities for grammatical and ungrammatical test items as a function of encoding condition. Like experimental participants, the HEM discriminated grammatical from ungrammatical test strings better in the trigram-encoding condition than in the bigram-encoding condition. However, the HEM makes this discrimination on the basis of a better hit rate and better correct rejection rate in the trigrams group, whereas the empirical data show that the locus of the better discrimination is mostly due to a better correct rejection rate alone. Although this mismatch does not impact the model's ability to predict the key discrimination, the imperfect fit is worth noting.

We used the model to explore the possibility that participants might have adopted other encoding schemes. For example, participants might have encoded strings as **a*b*1 + a*b*c*d*2 + a*b*c*d*e*f*3** (i.e., "*AB*, followed by *ABCD*, followed by *ABCDEF*") in the bigrams condition and

a*b*c*1 + a*b*c*d*e*f*2 (i.e., "*ABC* followed by *ABCDEF*") in the trigrams condition. Alternately, they might have encoded strings hierarchically as **a*b*1 + (a*b*1)* (c*d*2) + (a*b*1)*(c*d*2)*(e*f*3)** (i.e., "*AB*, *AB–CD*, and *AB–CD–EF*") in the bigrams condition and **a*b*c*1 + (a*b*c*1)*(d*e*f*2)** (i.e., "*ABC*, *ABC–DEF*") in the trigrams condition. Simulations with these encoding strategies failed to reproduce the experimental results, since neither set of alternative encoding schemes showed better discrimination in the trigrams than in the bigrams condition. We take those failures as corroborative evidence for our position that participants encoded strings as we had hoped that our experimental manipulation would make them. However, we acknowledge that there are other ways that participants might have encoded the strings in our experiment, and we invite researchers to adapt our code to examine their hypotheses about those other potential strategies.

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 follows from a process of retrospective inference from memory. 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 reported a new experiment to reinforce that performance in an artificial grammar task varies depending on how items are represented in memory. We have explained the data using a convolution-based model of memory that accommodates differential representations of items so that representations used in the simulation match participants' representations in the corresponding treatment conditions. The model we have proposed blends the holographic representation component of Jones and Mewhort's (2007) semantics model with the storage and retrieval components of Hintzman's (1986) MINERVA 2 model. Importantly, we wish to emphasize that the model fits reflect a priori predictions derived from the HEM before experimentation, rather than post hoc fitting. Given this, it is impressive that the model appreciates participants' overall better discrimination between rule-consistent and rule-violating items in Wright and Whittlesea's (1998) experiment (see Fig. 1) than of grammatical and ungrammatical items in our own experiment (see Fig. 4). We conclude that the holographic representation scheme of the HEM can be used to capture what people notice in training and test strings and that the exemplar-based model for storage and retrieval can be used to understand how exemplars are stored and deployed in the judgment-of-grammaticality task.

---

[3] Code: http://home.cc.umanitoba.ca/~jamiesor/LabPage/Data_models/Data_models.html.

Holographic representation has several strengths. First, convolution is a single method for encoding multiple forms of information: serial position information, serial order information, whole-item information, and hierarchical structure. Thus, convolution-based representation is a general and flexible framework for examining representation in memory and the influence that representation plays in performance and decision. Second, because convolution distributes over addition, it offers a distributed representation of exemplar structure that affords parallel access to stored information, thus solving the complications of computing the similarity between a probe and a memory trace that occur from feature misalignment in vectors under comparison. Third, holographic representation can be used to model several different cognitive problems—semantic representation (Jones & Mewhort, 2007), orthographic representation (Cox et al., 2011; Hannagan et al., 2011), logical inference (Eliasmith, 2004), free recall (Metcalfe-Eich, 1982; Murdock, 1982, 1983, 1995), memory disorder (Metcalfe, 1993), serial recall (Gabor, 1968, 1969; Longuet-Higgins, 1968; Poggio, 1973)—and so offers a general tool for representation over research domains. Finally, holographic representation is a biologically plausible model and so satisfies computational constraints imposed by brain physiology (Eliasmith, 2004).

Independently of our focus on artificial grammar learning, our analysis of representation and its impact on learning and memory in the artificial grammar task joins with a growing effort to criticize and improve representation in models of memory more broadly. A clear distinction can be drawn between the current convolution-based model and its concatenation-based predecessor (e.g., Jamieson & Mewhort, 2009, 2010). Because concatenation encodes serial position rather than serial order information in strings, it fails to capture the information that people extract and remember about training and test strings (i.e., groups of sequences or chunks). Convolution, by contrast, encodes serial order rather than serial position information and thereby successfully captures the information that people extract and remember about training and test strings. The distinction is at the heart of our position. We presented the same training and test strings to people in the bigrams- and trigrams-encoding conditions, but they behaved differently at test depending on how we instructed their encoding. If one cannot capture differences of representation, one cannot capture artificial grammar learning.

Our examination of representation in the artificial grammar task also joins a broader and emerging theme of recent research. Jones and Mewhort (2007) developed a computational model for learning the semantic structure of English. Johns and Jones (2010) used those structured representations to show that using random vectors to represent words limits the ability to make sense of peoples' performance in semantic memory tasks. Cox et al. (2011) made a similar argument by using holographic representation to show that modeling the orthographic structure in printed words is necessary to model peoples' performance in word recognition (see also Hannagan et al., 2011). The work here shows that building a better representation of letter strings helps to capture peoples' performance in an implicit learning task, an idea also explored by Dienes (1992), who compared performance in the artificial grammar task using single letters and bigrams in neural network models. Whereas a simplification of representation (i.e., use of random vectors) might have been a necessary strategy during early efforts to model memory retrieval, it is growing increasingly clear that representation is as important as retrieval and that a competent model of memory must attend to both problems.

Throughout this article and elsewhere (see Jamieson & Mewhort, 2005, 2009, 2010, 2011), we have argued for a retrospective account of performance in the artificial grammar task, in which participants infer the grammaticality of a test item by its similarity to memory for training exemplars. The position contrasts with the implicit learning view, in which the cognitive system abstracts the grammar prospectively in anticipation of future need. Redington and Chater (2002) previously underscored the same distinction in terms of lazy versus eager learning. Lazy learning is passive and involves storage of experiences and generalization of those experiences to new situations at test (e.g., judgments of grammaticality). Eager learning, on the other hand, is active and involves extraction of statistical regularities at study in service of unknown future goals.

We cannot help but see performance in the judgment-of-grammaticality and string completion tasks to be a lazy and retrospective process. At training, participants store the training items. At test, participants use their memory of the training items to answer the odd questions put to them. It is difficult for us to imagine how an eager learning system would know to extract a grammar in anticipation of something so unusual as a judgment-of-grammaticality task. Our computational model of performance shows that lazy and retrospective exemplar-based inference is sufficient to make sense of the empirical data.

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: Lawrence Erlbaum Associates 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.

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

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.

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: 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 & Development, 13,* 156–159.

Hannagan, T., Dupoux, E., & Christophe, A. (2011). Holographic string encoding. *Cognitive Science, 35,* 79–118.

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,* 126–137.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, and Computers, 16,* 96–101.

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

Jamieson, R. K., & Hauri, B. R. (2012). An exemplar model of performance in the artificial grammar task: Holographic representation. *Canadian Journal of Experimental Psychology, 66,* 98–105.

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.

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.

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.

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. (2011). Grammaticality is inferred from global similarity: A reply to Kinder (2010). *Quarterly Journal of Experimental Psychology, 64,* 209–216.

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

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

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.

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

Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review, 100,* 3–22.

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.

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 memory model for serial-order information. *Psychological Review, 90,* 316–338.

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

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.

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.

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.

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

Redington, M., & Chater, N. (2002). Knowledge representation and transfer in artificial grammar learning (AGL). In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making* (pp. 121–143). Hove: Psychology Press.

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

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.

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.