THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
iFirst, 1–26

Psychology Press
Taylor & Francis Group

Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity

Randall K. Jamieson
University of Manitoba, Winnipeg, Manitoba, Canada

D. J. K. Mewhort
Queen's University at Kingston, Kingston, Ontario, Canada

We present three artificial-grammar experiments. The first used position constraints, and the second used sequential constraints. The third varied both the amount of training and the degree of sequential constraint. Increasing both the amount of training and the redundancy of the grammar benefited participants' ability to infer grammatical status; nevertheless, they were unable to describe the grammar. We applied a multitrace model of memory to the task. The model used a global measure of similarity to assess the grammatical status of the probe and captured performance both in our experiments and in three classic studies from the literature. The model shows that retrieval is sensitive to structure in memory, even when individual exemplars are encoded sparsely. The work ties an understanding of performance in the artificial-grammar task to the principles used to understand performance in episodic-memory tasks.

Keywords: Artificial grammar; Global memory model; Implicit learning; Episodic memory.

People become sensitive to the regularities in a stimulus domain without deliberate effort and without explicit awareness. As a result, several theorists have argued for a specialized learning system that abstracts contingencies automatically and that guides behaviour adaptively.

Reber (1967) was the first to discuss the idea under the name *implicit learning*. His now classic experiment involved two groups of participants (see also Miller, 1958). Participants in a control group learned strings of letters assembled at random; participants in an experimental group learned strings constructed according to the rules of an artificial grammar. After they had learned the strings, Reber informed the participants that the strings had been constructed using rules and invited them to sort new grammatical test strings from ungrammatical ones. Participants from the experimental group sorted the two classes of stimuli better than the controls (achieving a score of 69% correct). Although participants could separate grammatical from ungrammatical strings, they could not explain the basis of their ability. In particular, they could not articulate the rules

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

The research was supported by grants from the Natural Sciences and Engineering Research Council of Canada to both the first and the second authors. We are deeply indebted to Beth Johns and Joaquin Vaquero for a critical reading and subsequent honing of the manuscript.

http://www.psypress.com/qjep
DOI:10.1080/17470210802055749

of the grammar. Reber argued that participants' ability to identify grammatical items reflected knowledge of the grammar but that, because the participants could not articulate the rules, the knowledge must be implicit. Following Reber's demonstration, others have endorsed the idea that participants abstract the grammar implicitly (e.g., Dienes, Broadbent, & Berry, 1991; Knowlton & Squire, 1992, 1993, 1994, 1996; Kuhn & Dienes, 2005; Manza & Reber, 1997; Mathews et al., 1989; McAndrews & Moscovitch, 1985; Reber, 1969, 1989, 1993; Rossnagel, 2001). The position that an implicit system operates independently of explicit intentional learning is known as the two-systems view. The two-systems view is attractive because it treats the artificial-grammar task as a laboratory demonstration of how people learn the rules of their native language without explicit awareness or effort. In Reber's (1967) words, abstraction in the artificial-grammar task is "a rudimentary inductive process that is intrinsic in such phenomena as language learning and pattern perception" (p. 863).

Despite its popularity, the idea of implicit learning remains controversial. Three kinds of argument have been marshalled against it. (a) The evidence that subjects abstract and use a grammar is almost always confounded with other possibilities (see Shanks & St. John, 1994). Instead of using the grammar, for example, participants might exploit information correlated with it, such as the similarity of the novel test exemplars to the known studied exemplars (e.g., Jamieson & Mewhort, 2005). (b) The evidence that learning is implicit—the fact that the participants cannot articulate the rules—is almost always incomplete. If their knowledge were probed appropriately, for example, participants might be able to articulate the grammar (e.g., Shanks & St. John, 1994, 1997). (c) The evidence for separate learning systems does not have the logical force needed to compel a two-systems view. Specifically, the dissociation between performance and awareness can be explained without invoking separate implicit and explicit learning systems (Dunn & Kirsner, 1988; Hintzman, 1990; Vokey & Higham, 2005).

The arguments against the implicit-abstraction idea have lead to alternative accounts of performance in the artificial-grammar paradigm. One position, called the *fragment view*, proposes that people parse the stimulus strings into grammatical primitives (e.g., pairs and triplets of symbols)—parsing is deliberate in some theories and automatic in others—and keep track of each primitive's frequency in the training materials. At test, strings that include a requisite number of the primitives, especially the most frequently occurring primitives, are judged to be grammatical (e.g., Johnstone & Shanks, 2001; Perruchet & Gallego, 1997; Perruchet & Pacteau, 1990; Perruchet & Vintner, 2002; Servan-Schreiber & Anderson, 1990). Thus, according to the fragment view, it is not necessary to learn the grammar per se; instead, participants keep track of grammatical primitives. Nonetheless, both the parsing operation and the definition of a primitive remain controversial. Some discuss parsing by analogy to a process that extracts word units from continuous speech (Perruchet & Pacton, 2005; Saffran, 2002). For others, parsing collapses to a count of the frequency with which symbols co-occur across training strings (e.g., Johnstone & Shanks, 2001; Knowlton & Squire, 1992, 1994; Perruchet & Pacteau, 1990).

A more extreme explanation for the artificial-grammar task, called the *exemplar view*, proposes that people infer the grammatical status of test strings by comparing them to memory of the training strings (see Brooks, 1978; Brooks & Vokey, 1991; Vokey & Brooks, 1992, 1994; Whittlesea & Dorken, 1993; Whittlesea & Wright, 1997; Wright & Whittlesea, 1998). Test strings that are similar to the training exemplars are taken to be grammatical; strings that are not similar to the training exemplars are taken to be ungrammatical (see Pothos & Bailey, 2000, for an application of Nosofsky's, 1991, generalized context model to the task). Because judgement of grammatical status is based on the comparison of each test string to memory, the exemplar view makes an understanding of retrieval central to an understanding of performance in the artificial-grammar task. A retrieval-driven explanation of grammaticality judgement stands in marked contrast to the prospective statistical-discovery process associated with the two-systems view. To date, however, there is little consensus about how memory of training exemplars drives performance in the artificial-grammar task. Further, the idea that people retain a perfect copy of the training exemplars seems to contradict the well-established limitation of memory (e.g., Cowan, 1995; Miller, 1956).

Although the implicit-learning view focuses on the distinction between performance and awareness, both the fragment view and the exemplar view sidestep the dissociation. Both deny that judgements of grammatical status depend on knowledge of the grammar, and if judgements of grammatical status do not depend on knowledge of the grammar, the dissociation presents no contradiction for theory to explain.

The grammar abstraction, fragment discovery, and exemplar-encoding views disagree about what information people acquire from the training exemplars. Because participants study the exemplars and not the grammar, however, all positions must agree that, whatever information is acquired, its source is the studied exemplars. At issue, then, is whether memory of the exemplars is sufficient to explain performance or whether extra information—such as implicit knowledge of the grammar or knowledge of statistical regularities in training materials—is required to understand judgements of grammatical status.

Because the exemplar-encoding view does not assume implicit knowledge, it does not need to postulate mechanisms to develop, store, or deploy implicit knowledge. Because it does not use a representation of aggregate structure in the exemplars (i.e., the grammar), auxiliary to memory for the exemplars themselves, the exemplar-encoding view is more economical than the two-systems position (e.g., Nosofsky & Zaki, 1998; Vokey & Brooks, 1992; Zaki & Nosofsky, 2001). Neal and Hesketh (1997) have reviewed a large body of work demonstrating a clear influence of exemplar knowledge in the artificial-grammar task. Jamieson and Mewhort (2005) provided recent support for the exemplar view. They quantified the structure in individual stimuli (a measure called local redundancy) and the structure in the grammatical rules from which the exemplars were derived (a measure called grammatical redundancy). The two kinds of redundancy are correlated; redundancy increases with grammatical redundancy. When separated experimentally, however, performance was predicted by local redundancy, not by grammatical redundancy. The aim of the current work is to develop that argument by describing a mechanism by which probe−exemplar comparisons drive performance in the artificial-grammar task.

The kind of grammar typically used in the artificial-grammar task is illustrated in Figure 1 (the grammar is taken from Dienes et al., 1991). Stimuli are generated by starting at the leftmost node marked 1 and by following paths until reaching an exit. 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 2, and 2 to 3 produces the string MTV. Exemplars produced using the grammar in Figure 1 vary on a number of dimensions including positional dependencies (e.g., strings can begin with only one or two symbols), sequential dependencies (e.g., each letter can be followed by only one or two others),

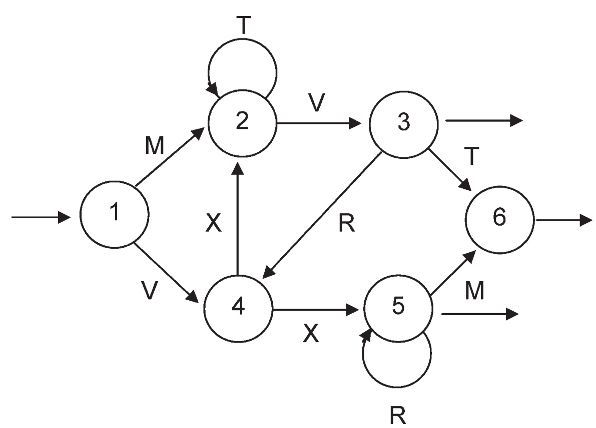

Figure 1. *A finite-state grammar. Grammatical strings are generated by entering the grammar at Node 1 (the leftmost node) and following paths (indicated by arrows) until an exit path is taken (the paths leading from Nodes 3, 5, and 6). When a path is taken, the associated letter is added to the end of the string. This grammar is taken from Dienes et al. (1991).*

string length, the frequency of repeated letters, and so forth. Because there are so many potential influences on performance, it is difficult to assess what information people use to judge a test string's grammatical status (see Johnstone & Shanks, 1999, for a clear review of the issue).

To model performance from experiments using such a complicated grammar, we would have to make a number of strong assumptions about the importance of the various factors and how they interact. For example, when a three-letter string is compared against a six-letter string, should the third letter of the three-letter string be compared against the final letter or to the letter in the corresponding position in the longer string? A model might fail because we have made the wrong assumptions. Alternatively, a model might provide a good fit to the data but the underlying account may hinge on risky assumptions. Because of the difficulties associated with such complex grammars, we generated three new experiments using stimuli that can be compared readily to each other. In Experiment 1, we used stimuli with structure defined by positional rules; such rules constrain only what symbols can occur in each position of a string. In Experiment 2, we used stimuli defined by sequential rules; such rules constrain what symbols follow one another in a string. In Experiment 3, we used stimuli defined by sequential rules and manipulated the amount of structure in the grammar. After we have reported the experiments, we will apply an exemplar-based theory of memory to each and then extend the model to examples of performance with traditional and more complex grammars. We will use the simulation to show that participants' ability to judge the grammatical status of test strings can be explained using principles that have long been applied to episodic memory.

EXPERIMENT 1: SENSITIVITY TO POSITIONAL DEPENDENCIES

In Experiment 1, participants viewed 20 grammatical training exemplars and were then asked to judge the grammatical status of 96 novel strings in a yes/no test procedure. To avoid contamination from previous experience, we constructed strings using nonalphanumeric symbols that participants were unlikely to be familiar with. To minimize the influence of any remaining idiosyncrasies in materials and to avoid item selection effects, we constructed a different stimulus set for each participant (see Murdock, 1982a; Redington & Chater, 1996). To avoid problems of comparability that occur with a complex grammar, we used a simple odd/even rule: Half the symbols occurred only in odd serial positions, and the other half only in even serial positions. To make the strings more directly comparable, we kept string length constant at four symbols. Finally, we manipulated the difficulty of the ungrammatical test items by varying the number of violations of the grammar. If the simple odd-even rule works like the complex grammar illustrated in Figure 1, participants should be able to infer grammatical status at above-chance levels without being able to articulate the simple generative rule. Finally, as the number of violations in a string increases, the probability that participants will endorse the string as grammatical should decrease.

Method

Participants

A total of 10 students from the Queen's University undergraduate participant pool took part in the study. All participants reported normal or corrected-to-normal vision. Each participant was tested individually.

Apparatus

The experiment was administered on a personal computer equipped with a 17-inch monitor and a standard mouse. Participants responded using the mouse to click on words that were displayed on the monitor. Stimulus strings were presented using a black 24-point font. At a viewing distance of 70 cm the visual angle of each symbol was approximately 1.64°. The background of the screen was light grey throughout the experiment.

Stimuli

The symbols used to construct strings were ♌, ♎, ♒, ♓, ♏, and ♐. For each participant, three of the symbols were assigned randomly to a set labelled *A*; the remaining symbols were assigned to a set labelled *B*.

Each string comprised four symbols and was constructed using an odd/even rule. According to the rule, only symbols from Set A could appear in Positions 1 and 3 of a string, and only symbols from Set B could appear in Positions 2 and 4 of a string. Ungrammatical items were produced by assigning one or more symbols to a position in a string in violation of the odd/even rule.

A unique stimulus set composed of 20 grammatical training strings and 96 new test strings was generated for each participant; 48 of the new test strings were grammatical, and 48 were ungrammatical. The ungrammatical strings varied according to the number of symbols that violated the odd/even rule: a total of 12 strings had one violation, 12 had two, 12 had three, and 12 had four violations. The position(s) of violations within each string was determined randomly. The test strings were presented in a random order.

Procedure

Participants were seated in front of the computer monitor. Before beginning the experiment proper, participants were given a chance to preview the six symbols. The participants were told that they would be shown 20 strings, that each comprised four symbols, and that their task was to remember the strings.

The participant initiated the training phase of the experiment by clicking on the message "I am ready to begin" displayed at the centre of the computer screen. The screen was cleared for 750 ms; then, the first training string was displayed at the centre of the screen. The string was displayed for 5 s; after the display, the screen was cleared for 750 ms. The cycle repeated until all 20 training strings had been presented. After all 20 strings has been presented, the participant was told that the strings had been constructed using rules, that an upcoming test phase would involve new strings, some constructed using the same rules and some constructed without the rules. Finally, the participant was given a new task—namely, to decide whether or not each new test string conformed to the rules.

On each test trial, a string of four symbols was presented at the centre of the screen, and the words "Consistent" and "Inconsistent" were displayed below it. The participant responded by clicking on the appropriate word. Once a response had been recorded, the screen was cleared for 500 ms; then the next test string appeared. The cycle continued until the participant had judged each of the 96 test strings (48 grammatical and 48 ungrammatical strings).

Finally, the participant was asked to state the rules that they thought had been used to construct the training materials and to describe any strategies that he or she used to judge the grammaticality of test probes.

Results and discussion

Grammatical probes were judged by participants to be consistent with the grammar 63% (SE = 2%) of the time; ungrammatical probes were judged by participants to be consistent with the grammar 48% (SE = 3%) of the time. By a signal detection analysis, discrimination was modest, $d' = 0.39$ ($SE = 0.1$); however, a single-sample t test confirmed that discrimination was better than chance, $t(9) = 4.12$, $\eta^2 = .65$, $p < .05$. The detection theory analysis also revealed a reliable bias on the part of participants to respond "consistent" ($C = -0.15$, $SE = 0.04$), $t(9) = 3.90$, $\eta^2 = .63$, $p < .05$.

The data indicate that the odd/even rule worked much like the complex grammars illustrated in Figure 1. Participants' ability to infer the grammatical status of the test probes is consistent with levels of performance from previous experiments using the judgement-of-grammaticality task (e.g., Dienes et al., 1991; Reber, 1967).

Figure 2 (closed circles) presents the percentages of test strings endorsed as grammatical as a function of the number of rule violations in the test strings. As is shown in the figure, the probability that an ungrammatical string would be

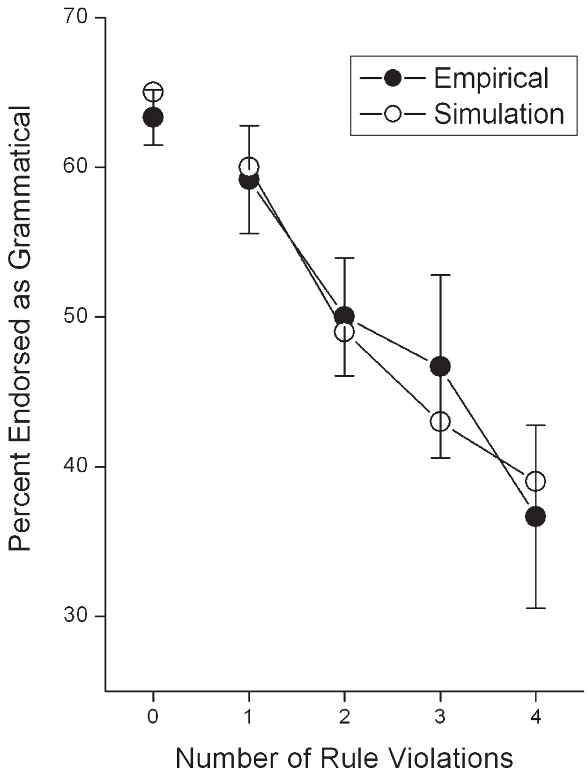

Figure 2. *Percentage of test strings endorsed as a function of the number of rule violations in Experiment 1. Strings with no rule violations were grammatical. Strings with 1 or more rule violations were ungrammatical. Closed circles show observed performance; open circles show simulated performance. Error bars represent standard errors. Parameters:* n = *20,* L = *.2,* k = *.39.*

endorsed as grammatical decreased linearly with the number of violations of the odd/even rule, $F(4, 36) = 5.33$, $MSE = 208.27$, $\eta^2 = .37$, $p < .05$; the linear trend accounted for a majority of the variance, $F(1, 9) = 13.38$, $MSE = 323.97$, $\eta^2 = .60$, $p < .05$.

Although the odd/even rule defines grammaticality by position, many of the resulting strings could have been classified on the basis of sequential dependencies: In grammatical strings, a character from Set A is followed always by a character from Set B and vice versa. Hence, the fact that performance is above chance does not indicate whether participants are sensitive to positional constraints or sequential constraints. The data for the ungrammatical strings disambiguate the question. The case of a *BABA* string contains four violations by a positional analysis but no violations to a sequential analysis. Because the four violation cases were rejected most easily, we infer that the participants were sensitive to the positional information. This inference should be interpreted with caution, however, because participants may have rejected strings with four violations because they had learned which symbols began and ended strings. In this case, the four violation cases would have contained two bigram violations in salient positions.

All of our participants expressed frustration when asked to articulate the rules used to construct the training items. When pressed, they reported selecting the string that "reminded them of" or that "looked most like" the training items. That is, they claimed that they relied on a test string's resemblance to the training materials in order to judge its grammaticality. When asked, none of our participants claimed to classify test strings according to positional or sequential dependencies of symbols in the training materials.

The results replicate the defining features from previous studies of the artificial-grammar task. The participants judged the grammatical status of test strings without being able to describe the rules, and judgements of grammatical status depended on the number of rule violations.

Although we replicated the standard results, the odd/even rule is not typical of grammars used in previous artificial-grammar experiments. Because strings were only four characters in length, the rule constrained severely the set of possible stimuli: Only 81 different grammatical strings of length four can be generated with the rule. Clearly, then, the grammar does not mimic the complexity of the grammars used in most studies using an artificial grammar. Secondly, finite-state grammars, like the grammar in Figure 1, emphasize sequential over positional dependencies. Because the first experiment used an atypical grammar, in Experiment 2 we used a grammar based on sequential dependencies that were uncontaminated by positional dependencies.

EXPERIMENT 2: SENSITIVITY TO SEQUENTIAL DEPENDENCIES

In Experiment 2, the symbols used to construct the strings were the digits 1 through 8. Unlike Experiment 1, all eight symbols were equally likely to appear as the first digit in a string. The grammar defined which digit could follow each other digit. That is, the grammar imposed sequential constraints on the strings without imposing a priori positional constraints. We used unfamiliar symbols to construct strings in the first experiment; in the second, we used familiar symbols (i.e., digits) and asked the participants to read the digits in a string aloud, from left to right. Reading the digits from left to right ensured that participants had ample opportunity to notice the sequential dependencies. The manipulation also exercised a degree of experimental control over encoding, a factor known to affect performance (Vokey & Brooks, 1992).

Participants studied 20 grammatical strings. Each string was composed of eight digits. After training, participants were informed that the strings had been constructed using rules and then were invited to sort novel grammatical from ungrammatical strings in a two-alternative forced-choice test. We used a two-alternative forced-choice test to remove distortions in performance from a bias in participants' predilection to respond yes or no more often in the yes/no procedure from Experiment 1 (see McAndrews & Moscovitch, 1985, for a discussion of the benefits of using a two-alternative forced-choice test). Each foil included at least one, and up to seven, violations of the grammatical rules. Thus, as in Experiment 1, we can examine performance as a function of the number of violations of the grammar. After the test, participants were asked to describe the rules that they believed were used to generate the training strings.

Method

Participants

A total of 10 students from the Queen's University psychology undergraduate participant pool participated in the study. All participants reported normal or corrected-to-normal vision. Participants were tested individually.

Stimuli

Stimuli were strings of eight digits. Only the digits 1 through 8 were used. The sequential rules constrained which digits could and which digits could not follow one another in successive serial positions. Table 1 illustrates the rules. The table gives the probabilities with which the digits 1 through 8 could follow one another in successive serial positions of a string. For example, when the digit 1 appeared at any position in a string, the digits 3, 5, 6, or 8 could follow, each with a probability of .25, but digits 1, 2, 4, or 7 could not occur.

Table 1 illustrates the rules in terms of how each digit can succeed one another. The same basic constraint was used for each participant,

Table 1. The grammar used to construct strings in Experiment 2

Digit at serial position n	Digit at serial position n + 1							
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1	0	0	1/4	0	1/4	1/4	0	1/4
2	1/4	0	0	1/4	0	0	1/4	1/4
3	0	1/4	0	1/4	1/4	0	1/4	0
4	1/4	0	1/4	0	0	1/4	0	1/4
5	0	1/4	0	1/4	0	1/4	1/4	0
6	1/4	0	1/4	0	1/4	0	0	1/4
7	1/4	1/4	1/4	0	0	1/4	0	0
8	0	1/4	0	1/4	1/4	0	1/4	0

but to guard against idiosyncratic digit pairings in strings (e.g., famous years or dependencies consistent with an odd/even dependency) each participant's grammar was determined by randomly reassigning digits to the row and column headings, subject to the constraints that (a) each digit could follow four other digits and could be followed by four other digits, and (b) digits could not follow themselves. As a result, each participant studied strings generated with a unique grammar that preserved the amount of constraint illustrated in Table 1.

The stimulus set constructed for each participant included 20 grammatical training strings, 50 novel grammatical test strings, and 50 ungrammatical test strings. Grammatical strings were constructed by selecting a digit at random for the first position in the string. Subsequent digits were added in accordance with the rules in the grammar until eight digits had been chosen. Ungrammatical strings were constructed by assigning a digit to each serial position of a string randomly, subject to the constraint that a digit could not follow itself. If the string contained at least one illegal bigram (i.e., if it included at least one pair of digits that violated the rules), it was used; otherwise, it was discarded. Sampling ungrammatical strings at random produced an uneven distribution: Participants were tested with a mean of 4.7, 10.6, 14.8, 10.6, 5.8, 2.8, and 0.7 strings with 1, 2, 3, 4, 5, 6, and 7 violations, respectively.

Apparatus
The apparatus was the same as that in Experiment 1.

Procedure
The experimenter instructed participants that they would be asked to read aloud strings of eight digits one digit at a time from left to right, that only the digits 1 through 8 would be used, and that each string would be presented individually for 10 s. The procedure differed from that in Experiment 1: We did not show participants the symbols used to construct strings (because participants were already familiar with digits 1 through 8).

The participant initiated the training phase of the experiment by clicking on the word "Study" displayed at the centre of the computer screen. The screen was cleared, and, after 750 ms, a string of eight digits was displayed at the centre of the screen. The participant read the string aloud. After 10 s of study time, the screen was cleared. After a 750-ms pause, the next string was displayed. Training continued until all 20 of the training strings had been presented.

After all the training strings had been presented, each participant was told that the strings had been constructed using rules and that they would be shown pairs of strings that they had not seen at study. Only one member of each pair conformed to the rules. Their task was to select the string that conformed to the rules used to construct the training items.

On each test trial, two unstudied strings, one grammatical and one ungrammatical, were displayed on the centre row of the computer screen. The position of the grammatical string (left versus right side) was determined randomly. The participant responded by clicking on the string that they thought was grammatical. As soon as a string had been selected, the screen was cleared. After a 500-ms pause, the next pair of test strings was presented. Testing continued until all 50 pairs of test strings had been presented, and the participant had selected one string from each pair. After completing the task, each participant was asked to describe the rules that had been used to construct the training strings.

Results and discussion

Participants selected grammatical test strings on 57% ($SE = 1\%$) of the test trials. Although performance was only 7% better than chance, a one-sample t test confirmed that performance was reliably better than chance, $t(9) = 6.70$, $p < .05$; the corresponding measure of effect size indicated that the difference was large, $\eta^2 = .83$. The reliable 7% advantage over chance is consistent both with the level of performance in Experiment 1 and with performance in other judgement-of-grammaticality experiments from the literature (e.g., Dienes, 1992; Reber, 1967).

Figure 3 (closed circles) shows the percentage of correct decisions as a function of the number of rule violations. The degree of variation in the size of the standard errors in Figure 3 reflects differences in the number of test trials on which participants were presented an ungrammatical string with 1, 2, 3, 4, 5, 6, and 7 rule violations. The relation between performance and number of rule violations in the current experiment (Figure 3) is consistent with the relation observed in Experiment 1 (Figure 2): Judgement of grammatical status improved with the number of violations in the ungrammatical alternative.

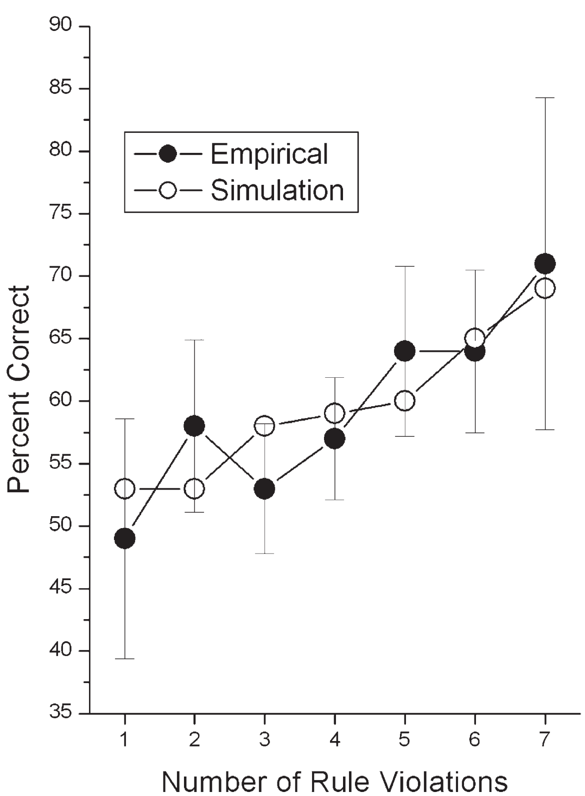

Figure 3. *Percentage correct in Experiment 2 as a function of number of rule violations in the ungrammatical alternative. Closed circles show observed performance; open circles show simulated performance. Error bars represent standard errors. Parameters:* n = *20,* L = *.55.*

All 10 of our participants were unable to articulate the structure of the training materials. None of our participants felt they could provide a rule that would help a yoked participant judge the grammaticality of the test strings they had just viewed. When pressed, participants said they relied on their "gut feeling"; none described specific sequential dependencies or fragments (i.e., groupings of digits) that they might have treated as grammatical primitives. A total of 8 of the 10 participants expressed doubt that the training materials were constructed using rules, thinking that deception was part of the procedure.

The experiment replicates the classic results from earlier experiments using the judgement-of-grammaticality procedure: Participants inferred the grammatical status even though they were unable to report the rules of the grammar. Whereas in Experiment 1 participants were sensitive to positional constraint of symbols, in Experiment 2 they were sensitive to sequential constraint of symbols.

EXPERIMENT 3: AMOUNT OF STRUCTURE

In both Experiments 1 and 2, performance varied with the number of grammar violations. However, a count of violations is a primitive measure of structure, and it is theoretically loaded. In Experiment 3, we used the measure of grammatical structure developed by Jamieson and Mewhort (2005) to manipulate the amount of constraint in the grammars that were used to generate study strings. In addition, we manipulated the number of training exemplars, either 20 or 40 digit strings. Both factors manipulate the information available for learning and so provide better grounds on which to analyse learning. The procedure was identical to that from Experiment 2: After reading the training strings, the participants were told that the strings had been constructed using rules and were invited to discriminate novel grammatical probes from ungrammatical probes.

Method

Participants

A total of 56 students from the McMaster Psychology undergraduate participant pool took part in the study as part of a course requirement. Participants were assigned to one of eight experimental conditions defined by the factorial combination of amount of structure and the number of exemplars presented in training. All participants reported normal or corrected-to-normal vision; participants were tested individually.

Apparatus

The apparatus was the same as that used in the previous experiments.

Stimuli

Stimuli were strings of eight digits. Only the digits 1 through 8 were used.

Four grammars that differed in grammatical redundancy were used to construct the stimuli. In the least constrained condition, each digit could be followed by six other digits with equal likelihood—that is, $p = 1/6$. The next condition was the same as Experiment 2: Each digit could be followed only by four other digits with equal likelihood—that is, $p = 1/4$. In the next condition, each digit could be followed only by three other digits, again with equal likelihood—that is, $p = 1/3$. Finally, in the most heavily constrained grammar, each digit could be followed only by two other digits with equal likelihood—that is, $p = 1/2$. The amount of structure associated with each grammar was quantified using the redundancy statistic, G (see Jamieson & Mewhort, 2005). Redundancy is computed using Shannon and Weaver's (1949) expression for uncertainty:

$$U = -\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}\, log_2\, p_{ij},$$

where p_{ij} denotes the probability of symbol j following symbol i in a sequence, and n is the number of symbols in the grammar. We compute the grammatical redundancy of a target grammar by comparing its uncertainty, *U(Grammar)*, against the uncertainty in an equivalent but unconstrained grammar, *U(Unconstrained)*:

$$G = 1 - \frac{U(Grammar)}{U(Unconstrained)}$$

where G measures the amount of bigram structure associated with a grammar. Using Table 1 as an example, an unconstrained grammar would allow each symbol to follow each other symbol equally often (each entry in the table would be 1/8). Moving from the least to the most constrained grammars used in the current experiment, $G =$ .14, .33, .47, and .67.

As in Experiment 2, to construct a lure, digits were selected for each position of a string at random, and the string was evaluated as to whether it violated the rules of the grammar. The probability of obtaining a violation is related to grammatical redundancy: The more constrained the grammar, the easier it is to introduce a violation by chance. We did not anticipate the potential confound, but offer an analysis of it following the experiment. Further, as in Experiment 2, the particular stimuli used by each participant were determined by assigning randomly digits to the row and column headings of the transition table that defined the grammar.

Procedure

Training and test procedures were identical to those from Experiment 2, except that half of the participants read aloud 40 rather than 20 training exemplars.

Results and discussion

Figure 4 (top panel) shows the percentage of trials on which participants selected a grammatical test string as a function of grammatical redundancy and number of training exemplars. Single-sample t tests confirmed that discrimination was better than chance (50%) in all but two conditions ($G = .33$ and .14, both with 20 training exemplars).

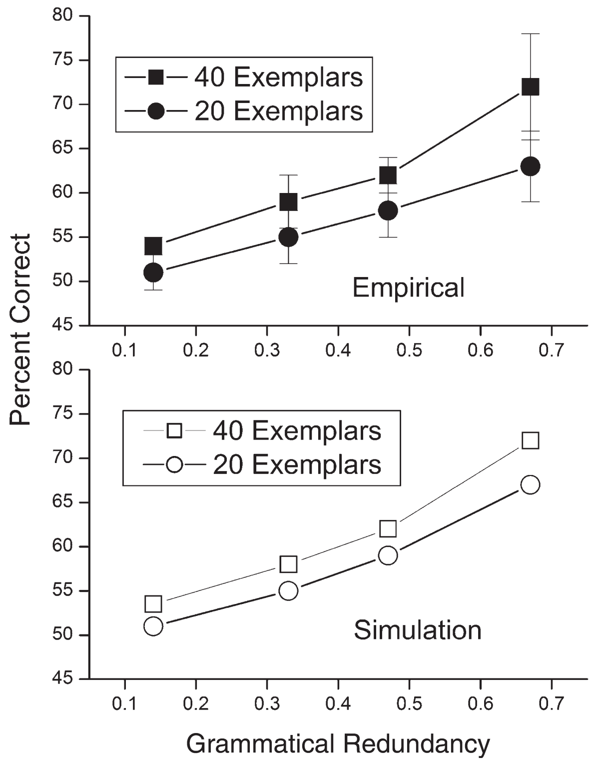

Figure 4. *The top panel shows percentage correct in Experiment 3 as a function of grammatical redundancy and the number of training exemplars. Error bars represent standard errors. The bottom panel shows simulations of Experiment 3. Parameters:* n = *20,* L = *.45.*

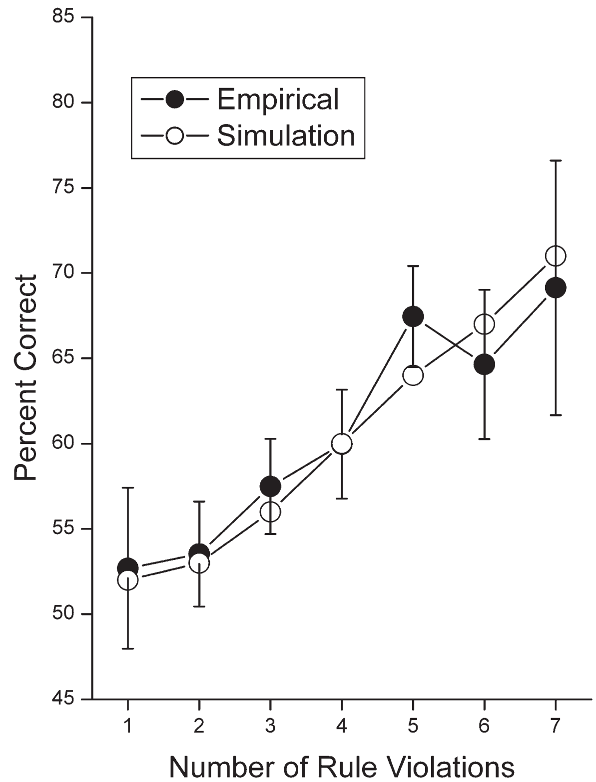

Figure 5. *Percentage correct in Experiment 3 as a function of number of rule violations in the ungrammatical alternative. Closed circles show observed performance; open circles show simulated performance. Performance is collapsed across all eight cells in the experimental design. Error bars represent standard errors. Parameters:* n = *20,* L = *.45.*

Correct choices increased linearly from 52% (SE = 2%) to 68% (SE = 5%) as a function of grammatical redundancy, $F(1, 48) = 23.74$, $MSE = 69.41$, $\eta^2 = .30$, $p < .05$, and accuracy was 5% better after training with 40 strings than after training with 20 strings, $F(1, 48) = 5.48$, $MSE = 69.41$, $\eta^2 = .07$, $p < .05$. There was no evidence that extended practice interacted with grammatical redundancy, $F(3, 48) = 0.24$, $MSE = 69.41$, $\eta^2 = .01$, $p > .05$.

As in Experiments 1 and 2, participants improved at rejecting strings as the number of rule violations increased (see the closed circles in Figure 5). Participants were tested with a mean of 9.6, 7.8, 8.7, 9.1, 8.3, 5.0, and 1.5 strings with 1, 2, 3, 4, 5, 6, and 7 violations, respectively. The variation in the size of the standard errors in Figure 5 is exaggerated by differences in the number of test trials on which participants were presented an ungrammatical string with 1, 2, 3, 4, 5, 6, and 7 rule violations.

Following the test phase, we asked participants whether they could articulate the rules of the grammar. None of our participants articulated the rules. Next, we asked participants to provide a rule that we could give to a yoked participant so that the yoked participant would be able to reproduce their decisions: Participants were unanimously reluctant to give the verbal rule, and many refused. When pressed, some of our participants gave rules that identified specific groups of digits that they remembered from the training materials (e.g., "Some of the strings had a 1 and an 8, and sometimes a 5 and 2 were beside each other"). However, participants who gave these rules expressed very little confidence in the rules they gave. The 3 participants who claimed their rules would help a yoked participant asserted that the rule they provided would only work some of the time.

The experiment replicated the classic results from earlier experiments. It also showed that discrimination improved as a function of both amount of structure and the number of training exemplars. Finally, although participants were able to assess the grammatical status of the test strings, they were unable to describe the grammar. Instead, they reported strategies based on the similarity of the test strings to the training strings.

The fact that participants in all three experiments were able to exploit structure is not surprising: People exploit structure in many situations. The surprise is that participants were sensitive to subtle contingencies after so little exposure to them, even when they were unaware that the contingencies were available to be learned.

Although the exemplar view assumes only that participants encode the exemplars, encoding multiple strings presents its own difficulties. Memory is limited. The limitation is usually discussed in terms of the number of whole items that people can encode and hold in short-term store (e.g., Miller, 1956; see also Cowan, 1995). Instead of treating memory as a store with a fixed capacity measured in whole items, perhaps a realistic alternative is that participants encode large numbers of items but that encoding is both noisy and imperfect. In the following section, we use the imperfect-encoding idea to show how an exemplar model of memory uses a noisy representation of the exemplars to drive performance.

UNDERSTANDING THE ARTIFICIAL-GRAMMAR TASK

When Reber (1967) reported his classic experiments, influential theories of concept formation (developed by Bruner, Goodnow, & Austin, 1956) proposed that category learning was as a deliberate process of induction: Participants parsed stimuli into bundles of features and developed rules that related the features to the categories. An implicit learning system that performed the inductive process independently of awareness was a natural extension of existing wisdom.

In the next two decades, work in memory and categorization grew in sophistication, and theorists developed ways to represent stimuli and to describe the processes that operate on them. In current models of episodic recognition, for example, encoded items are represented in a graded rather than all-or-none fashion; that is, items are encoded imperfectly to different degrees of resolution (e.g., Hintzman, 1984). Some models even merge information about separate items into a single holographic structure (Eich, 1982; Murdock, 1982b, 1983). To replace the serial-search idea, techniques were developed for comparing a single item against a set of items. The concept of parallel retrieval encouraged the development of the global memory models (e.g., Eich, 1982; Gillund & Shiffrin, 1984; Hintzman, 1984; Murdock, 1982b, 1983; Pike, 1984; Raaijmakers & Shiffrin, 1981; see Clark & Gronlund, 1996, for a review). Parallel access was also used in models of item categorization, where exemplars are represented as points in a multidimensional psychological space (Medin & Schaffer, 1978), and categorization is determined by computing the sum of the distances of a probe to all previously encountered exemplars from the category in question; the smaller the distance, the more likely the probe is to be a member of the category (see also Nosofsky, 1984, 1986, 1991; Nosofsky & Palmeri, 1997).

In their work, Brooks and his colleagues have promoted the principles of an exemplar account for understanding judgement of grammaticality, but have left the development of a working model as a promissory note (e.g., Brooks, 1978; Brooks & Vokey, 1991; Vokey & Brooks, 1992). Here, we fulfil their promise by adapting a multitrace model of memory to the task of judging grammatical status.

Minerva 2

Minerva 2 (Hintzman, 1984, 1986, 1988, 1990) is a multitrace model of memory. The model was developed initially to understand the episodic-recognition and judgement-of-frequency tasks. It has since been applied to a wide range of phenomena including categorization (Hintzman, 1986, 1988), confidence–accuracy inversions in recognition memory (Clark, 1997), recognition failure of recallable words (Hintzman, 1987), false recognition in the Deese–Roediger–McDermott (DRM) paradigm (Arndt & Hirshman, 1998), clinical diagnosis (Dougherty, Gettys, & Ogden, 1999), extrapolation in function learning (Kwantes & Neal, 2006), speech perception (Goldinger, 1998), word naming (Kwantes & Mewhort, 1999), and access to semantic memory (Kwantes, 2005).

Minerva 2 proposes that when a participant encounters an item, the item is encoded to memory as a separate trace. When a recognition probe is presented, it is compared, in parallel, to all traces in memory, and each trace is activated in proportion to its similarity to the probe. The activated exemplars are merged into an *echo*. The echo's *content* is an aggregate of the information in memory activated by the probe. The echo's *intensity* measures the activation triggered by the probe and is often used as the basis for judgements of familiarity.

In the model, a stimulus is represented by a vector of n elements. Each element takes one of two values: $+1$ or -1 with equal probability—that is, $p(+1) = p(-1) = .5$. An association between two stimuli (or between a stimulus and a response, or a stimulus and a category label) is represented by concatenating the constituent item vectors to form a new vector of double dimensionality.

In Minerva 2, memory is a matrix with one row (vector) for each studied event. Encoding an event involves copying its vector to a new row in the memory matrix. Encoding can be imperfect. The model accommodates variation in the quality of encoding by varying the number of elements in a stimulus vector that are stored correctly. If a particular element is not stored correctly, its value is set to 0 (indicating that it is indeterminate or unknown). The parameter L controls the probability with which an element is stored. As L increases, the resolution of the encoded stimulus improves. Minerva 2 treats forgetting as the inverse of correct encoding; hence, L is used also to accommodate memory loss.

In the model, all retrieval is cued. When a cue is presented, it activates all memory traces in proportion to their similarity to the cue. The activation from all traces is aggregated into a composite trace (the echo). Similarity of trace, i, to the probe, P, is given by

$$S_i = \frac{\sum_{j=1}^{n} P_j \times M_{ij}}{n},$$

where P_j is the value of the jth feature in the probe, M_{ij} is the value of jth feature of the ith row in memory, and n is the number of features in the vectors under comparison. Like the Pearson r, the similarity of the ith item to the probe, S_i, is scaled to the interval $\{-1, +1\}$ by dividing the numerator (the dot-product) by n. Similarity equals $+1$ when the row is identical to the probe.

The ith trace's activation, A_i, is the cube of the similarity to the probe,

$$A_i = S_i^3.$$

The activation function exaggerates the differences in similarity between the probe and the items in memory by attenuating retrieval of exemplars that are dissimilar or only moderately similar to the probe. Note that using an odd-numbered exponent in the activation function preserves the sign of the argument, S_i.

The echo, C, is a vector obtained by weighting each of the $i = 1 \ldots m$ traces in memory by its activation and summing all m traces into a composite,

$$C_j = \sum_{i=1}^{m} A_i \times M_{ij}.$$

The echo represents the amalgam of traces that the probe retrieves from memory and is used to model experiments in which information must be recovered from memory.

The echo's intensity is computed by summing activation across the m traces in memory—that is, $i = 1 \ldots m$,

$$I = \sum_{i=1}^{m} A_i.$$

The echo's intensity, I, ranges between $-m$ and m, and it is a function of both the number of traces in memory and the degree to which the traces match the probe. If all traces in memory were identical to the probe, $I = m$. Echo intensity is used often as an index of the probe's similarity to items in memory. However, because the intensity measure is additive, it can vary dramatically with the number of items stored in memory. An alternative measure of intensity is obtained by computing the similarity between normalizing the echo and it and the probe (see Hintzman, 1988, p. 546):

$$I = \frac{\sum_{j=1}^{n} P_j \times C_j}{n},$$

where P_j is the value of the jth feature in the probe, C_j is the value of jth feature in the echo, and n is the number of features in the vectors under comparison. Because we compare performance with different numbers of training items, we adopt the latter method.

Adapting Minerva 2 to Experiment 1

In Experiment 1, the strings comprised four unfamiliar characters organized according to an odd−even positional rule. To implement the rule, the pool of six characters was divided into two sets; the characters from one set were assigned odd positions, and the characters from the other set were assigned even positions.

To simulate the task, we started by constructing six vectors, one to stand for each character. Each character vector was of dimensionality 20 with values of $+1$ or -1 selected at random with $p(+1) = p(-1) = .5$. When participants studied the strings, we assumed that they treated each string as a four-character unit. Accordingly, we represented each string in memory by concatenating four character vectors to form a single vector composed of four successive subfields, one subfield for each symbol in the string.

Each participant studied 20 training strings. Accordingly, we stocked memory with 20 vectors, each vector corresponding to one of the 20 studied strings. Hence, after encoding, memory comprised 20 vectors of 80 elements.

The stimulus strings were unfamiliar characters; hence, it is unlikely that participants encoded them well. Accordingly, we set the learning parameter to a low value, $L = .2$. As a result, about 80% of the elements in each string were indeterminate, representing sparse encoding of items.

To simulate the judgement-of-grammaticality task, we constructed vectors to represent the probes used in Experiment 1; the probes were constructed in the same fashion as the study strings. Each probe was applied to memory, and an echo was produced. A string was endorsed as grammatical if the echo intensity exceeded a decision criterion, $k = .39$. To obtain stable point estimates of performance, we averaged across 100 independent simulations of the experimental procedure.

Simulated performance matched participants' performance in Experiment 1 closely. In Experiment 1, the hit and false-alarm rates were .63 ($SE = .02$) and .48 ($SE = .03$), respectively, for a $d' = 0.38$. By comparison, the simulated hit and false-alarm rates were .65 and .47, respectively, for a $d' = 0.45$. The model fit yielded a bias toward a "consistent" response, $C = -0.16$; the same bias observed in the corresponding experiment, $C = -0.15$.

Simulated performance also matched the typicality gradient of the empirical data. The open circles in Figure 2 show the percentage of test strings identified as grammatical in the simulation as a function of the number of rule violations in test strings. As is clear in Figure 2, means computed from the simulated data provided a close match to the means computed from participants' performance: The greater the number of rule violations in a test string, the less likely that the string was endorsed, $RMSE = 2.53\%$.

The typicality gradient is easy to understand in terms of the global similarity measure. A grammatical test string never matched a single training string on all four symbols, but it matched several training strings on one to three symbols. Each additional rule violation in an ungrammatical string decreased the number of symbols that matched individual training items and, thereby, decreased the string's global similarity to the training set.

The simulation exhibits the critical features from studies of implicit learning using the artificial-grammar task. The model does not know the grammar, either explicitly or implicitly, yet it captures judgements of grammaticality. We take the demonstration as evidence that it is not necessary to suppose that participants learned and applied implicit knowledge of the grammar in order to understand their performance. Instead, we conclude that performance reflects well-known principles of memory, principles that have been exploited to understand performance in recognition and classification experiments for more than 20 years.

The exemplar view differs from the implicit-learning view in four main ways. (a) The implicit-learning view claims that the regularities in training strings are available in a centralized form, likely as a set of rules. In the exemplar view, the rules used to construct the training strings are not represented in memory, but information about the regularities produced by the rules is represented indirectly in the encoded exemplars. People learn the exemplars; they do not learn the grammar. (b) According to the implicit-learning view, participants abstract the grammar during training and then apply it during test. On the exemplar view, participants store exemplars and judge a probe's grammatical status using its similarity to known grammatical cases. (c) In the implicit-learning view, knowledge about the grammar is compiled *prospectively* in the form of grammatical rules. In the exemplar view, judgements of grammaticality are computed *retrospectively* at the time of test by a comparison of the probe to memory of studied items. (d) The implicit-learning view supposes that performance in the artificial-grammar task reflects mechanisms in a specialized learning system. The exemplar view attributes performance to the same mechanisms as those that are used widely to understand performance in episodic-memory tasks.

Adapting Minerva 2 to Experiment 2

In Experiment 2 we tested learning of sequential dependencies. The stimuli were strings of eight digits constructed according to sequential rules. To simulate the task, we started by constructing eight unique 20-element vectors, one for each of the eight digits. The vectors were concatenated to form 20 training strings of 160 elements composed of eight successive subfields, one subfield for each position in the string, and 100 novel test strings of 160 elements (50 grammatical and 50 ungrammatical).

We stocked memory with the training strings, 1 row per string. Hence, after encoding the training strings, the memory matrix had 20 rows (1 for each training string) and 160 columns. When we simulated Experiment 1, we set the learning parameter very low to acknowledge the difficulty that participants had when encoding unfamiliar characters. Because the stimuli in Experiment 2 were familiar digits, we set the learning rate higher, $L = .55$. Nevertheless, the matrix was still sparse with almost half of its elements set to zero.

Each test involved one grammatical and one ungrammatical test string. Both test strings were compared against memory and generated two separate echoes. The test string with the greater echo intensity was selected as the grammatical alternative. We conducted 100 independent replications of the procedure to obtain stable point estimates of performance. In the simulation, grammatical test strings were selected on 57% of the trials matching the score of 57% correct ($SE = 1\%$) obtained in Experiment 2.

The open circles in Figure 3 show simulated percentage correct as a function of the number of rule violations in the ungrammatical alternative of each test pair. For both the model (open circles) and the participants (closed circles), the probability of choosing the grammatical string over its ungrammatical alternative increased with the number of violations in the ungrammatical alternative. The correspondence between simulated and observed performance shown in Figure 3 was high, $RMSE = 3.74\%$.

The model accommodated the results from Experiments 1 and 2—materials constructed according to positional and sequential dependencies, respectively—without requiring modification of its encoding and retrieval assumptions. In fact, we removed a free parameter (i.e., the criterion) by using a two-alternative forced-choice test procedure. The success of our simulations underscores our main point: One can assess the grammatical status of a test item based on its global similarity to studied exemplars and without knowledge of the rules used to construct training items; one does not need to know the grammatical rules to show sensitivity to contingencies in studied exemplars. Performance in the judgement-of-grammaticality task is explained using the same mechanisms as those that are used widely to understand phenomena of episodic memory.

Adapting Minerva 2 to Experiment 3

Next, we applied the model to the factorial design in Experiment 3. Except for variation in both grammatical redundancy and the number of training exemplars, the simulations were the same as those for Experiment 2. Because participants received less study time per item in Experiment 3 (8 s per item) than in Experiment 2 (10 s per item), L was lowered to .45.

Figure 4 (bottom panel) shows the percentage of trials on which a grammatical string was chosen correctly as a function of both grammatical redundancy and the number of training exemplars. As in the observed data (see Figure 4, top panel), discrimination improved as a function of both factors. The cell-by-cell correspondence between simulated and observed performance was high, $RMSE = 1.78\%$.

Figure 5 shows performance as a function of the number of rule violations in the ungrammatical alternative. As is clear in the figure, the simulated typicality gradient matched the empirical one, $RMSE = 4.83\%$.

The model captured performance as a function of both grammatical redundancy and the number of exemplars presented in training. Neither result is surprising from the perspective of episodic-memory theory. First, by increasing the number of exemplars, the strings stored in memory provide a better representation of the contingencies derived from the grammar; measurement of echo intensity from a larger N is influenced by outliers less than it is from a small N (the central limit theorem). Second, increasing grammatical redundancy makes the exemplars that are produced by the grammar more homogeneous. As a result, increasing redundancy of the grammar forces grammatical and ungrammatical probes to be more and less like the exemplars stored in memory, respectively.

To illustrate this point, we present Figure 6. Figure 6 illustrates the shift in similarity for test

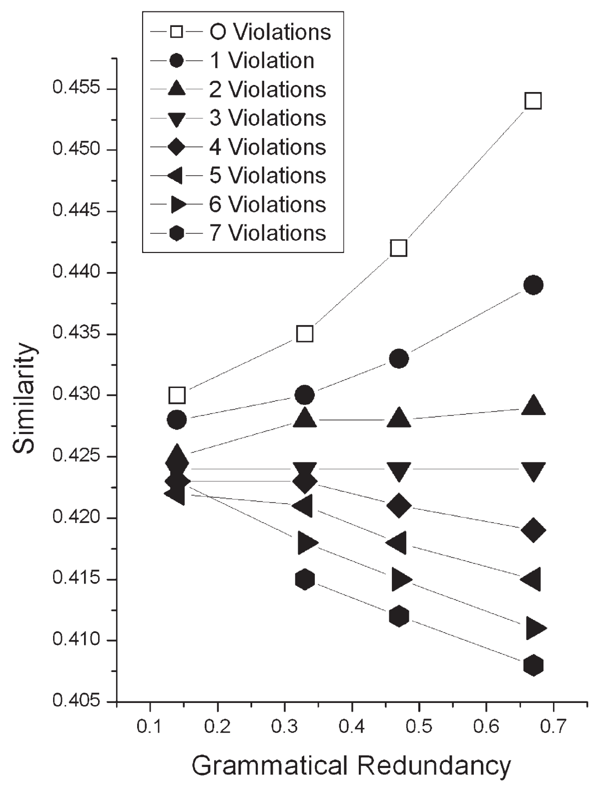

Figure 6. *Similarity of the probe to the echo as a function of grammatical redundancy and number of rule violations. Curves correspond to the 40-exemplar condition.*

probes as a function of both grammatical redundancy and number of rule violations. The data are taken from the simulation for the 40-exemplar cell. As shown, echo intensity of grammatical probes (i.e., zero violation probes) increased as a function of grammatical redundancy. For ungrammatical strings, the slope of echo intensity as a function of grammatical redundancy decreased the greater the number of grammatical violations in a string; for ungrammatical strings with more than four violations, the slope was negative. These two factors combined to produce increasingly better discrimination of grammatical and ungrammatical probes as grammatical redundancy was increased. The pattern of performance in the bottom panel of Figure 4 is a direct consequence of this factor and accounts for participants' performance in the experiment shown in the top panel of Figure 4.

Adapting Minerva 2 to complex grammars

Most work with the artificial-grammar task has used complex grammars of the sort illustrated in Figure 1. Having shown that the model captures performance in experiments with contingencies based on simple position and sequential rules, we turn our attention to experiments with contingencies based on complex grammars.

Reber (1967). Reber's (1967) seminal study has served as the point of departure for most modern work with the artificial-grammar task. In his study, participants studied grammatical exemplars and, after training, were invited to judge the grammatical status of novel test probes. Reber's participants (Exp. 2) achieved 69.4% correct, but were unable to articulate the grammar.

We extended the model to deal with Reber's (1967) materials. We started by constructing five unique 20-element vectors, one for each of the five letters that Reber used to construct his materials: *P*, *S*, *T*, *V*, and *X*. In the experiment, participants studied 20 grammatical strings and then were tested with 24 novel grammatical test strings and 24 ungrammatical test strings. Unfortunately, Reber did not list the specific study items in the 1967 paper. He did, however, list a representative example of strings from his grammar elsewhere (Reber, 1993, p. 36). We took our strings (both grammatical and ungrammatical) from the latter source that comprised a study list of 20 grammatical training strings, 25 grammatical test strings (6 of which were in the set of 20 training strings), and 25 ungrammatical test strings.

The letter-vectors were concatenated to form 70 string-vectors of 160 elements corresponding to the training and test strings. As we noted earlier, because strings varied in length, we were forced to invent a way for comparing strings of different length. We elected to compare the letters of the probe to the letter in the corresponding position each string in memory. Accordingly, we treated each stimulus as a string with eight letters. For strings with fewer than eight letters, a vector of zeros coded each blank position. As a result, the string *MTV* (a popular TV channel not available at the time of the original experiment) was encoded as a string with eight letters but the elements in the fourth to last subfields of the vector were set to zero. We stocked memory with the 20 training strings, one row per string. After encoding the training strings, the memory matrix had 20 rows (one for each training string) and 160 columns.

To simulate Experiments 1–3 reported in this paper, we set the learning parameter to a low value to acknowledge the difficulty of encoding briefly presented strings. In addition, we had participants read the strings one at a time. By contrast, Reber (1967) presented sets of four strings and required the participants to reproduce the set twice before moving onto a new set of strings. As a result, his participants must have encoded the materials more completely than our participants did. Nevertheless, encoding cannot have been perfect, because successive sets would likely interfere with memory for the earlier sets. To acknowledge the difference between our experiments and Reber's, we set the learning parameter to a much larger value than before, $L = .75$. Accordingly, after encoding the strings, memory for individual strings was much stronger.

As before, each test string was applied to memory, and an echo was produced. The test string was endorsed as grammatical if the echo intensity exceeded a decision criterion. Because we assumed better encoding with Reber's (1967) procedure than with ours, memory was less sparse than that in our earlier simulations, and we had to rescale the decision criterion. Using a higher criterion, $k = .63$, the model matched performance of Reber's (1967, Exp. 2) participants, 69% correct: Simulated hit and false-alarm rates were .77 and .39, respectively. A model developed to explain recognition and categorization performance can explain performance in the judgement-of-grammaticality task.

Dienes, Broadbent, and Berry (1991). Dienes et al. (1991; Exp. 1) manipulated the grammatical homogeneity of the study materials in an artificial-grammar task. Participants in one group, the *pure group*, studied 20 grammatical training strings before judging the grammatical status of 25 grammatical and 25 ungrammatical test probes in a yes/no test procedure. Participants in the second group, the *mixed group* studied 20 grammatical and 20 ungrammatical training strings before judging the grammatical status of 25 grammatical and 25 ungrammatical test probes in a yes/no test procedure. Participants in the pure group achieved 65% correct ($SE = 2\%$); those in the mixed group achieved 60% correct ($SE = 1\%$).

From the current model's perspective, the mixed group presents a challenge. When the participants in the mixed group studied the nongrammatical strings, the fact that the strings were not grammatical was correlated with colour of the display. The colour may have allowed the participants to distinguish grammatical from ungrammatical strings at study; even though they probably had neither motivation nor explicit instructions to do so. The model does not include a mechanism with which to notice the signal; instead, all of the exemplars—both grammatical and nongrammatical—are treated as if they were grammatical. Hence, memory includes as many ungrammatical strings as grammatical strings.

Although half the studied items were ungrammatical, the manipulation of grammatical status is unlikely to have destroyed the pattern of contingencies established by the grammatical items. Violations of the grammar were introduced by "substituting an inappropriate letter for an appropriate letter in an otherwise grammatical string. The position of the violation covered letter positions one to six of the 15 exemplars" (Dienes et al., 1991, p. 877). The effect of the manipulation should be to weaken the contingencies aligned with the offending letter, but to strengthen the contingencies aligned with the remaining letters. At issue, then, is how the two forces balance. It seems safe to anticipate that the pure group will represent the contingencies of the grammar more accurately than will the mixed group. For that reason, performance should be better in the pure group than in the mixed group.

The materials are provided in Dienes et al.'s (1991) Table 1. The strings were constructed from the consonants *M*, *T*, *V*, *R*, and *X* and included between three and six consonants. Each training string was studied for 5 s. Five of the grammatical training strings and five of the ungrammatical training strings were presented both at study and at test.

To adapt the model to Dienes et al.'s (1991) materials, we started by constructing five unique 20-element vectors, one for each of the five letters used to construct the strings: *M*, *T*, *V*, *X*, and *R*. The vectors were concatenated to form 90 vectors of 120 elements that corresponded to each training and test string (see Appendix for a different representation scheme used by Dienes, 1992). To simulate the pure group, we stocked memory with the 20 grammatical training strings, one row per string. To simulate the mixed group, we stocked memory with the 20 grammatical and 20 ungrammatical training strings. Hence, for the pure group, the memory matrix had 20 rows (1 for each training string) and 120 columns; for the mixed group, the memory matrix had 40 rows and 120 columns.

Each of the 50 test strings was applied to memory, and an echo was produced for each. A string was endorsed as grammatical if the echo intensity exceeded a decision criterion. We set L to .35 to reflect poor encoding of stimuli, and we set k to .45.

Simulated performance of the pure and mixed groups—65% and 59%, respectively—closely matched performance of Dienes et al.'s (1991) participants in the corresponding conditions—65% and 60%, respectively. As anticipated, the pure condition represented the contingencies of the grammar more accurately than did the mixed condition. A violation of the grammar produced by substituting an incorrect item for a correct one is a local perturbation; it weakens the contingencies aligned with the offending item, but the rest of the string strengthens the rest of the grammar's contingencies. The decision is based on a global measure of the match of the probe to memory. The global measure is able to balance the local effect of a violation against the broader effect of strengthening valid contingencies in memory. In a rule-based system, by contrast, if ungrammatical strings were treated as valid items, the ungrammatical items would contradict the rules, and, for that reason, a rule-based system is likely to become confused. The principle at work here is that a similarity-based model balances the partial contradiction; it is more accommodating to perturbations in regularity than a system in which a single rule violation invalidates a whole string.

Vokey and Brooks (1992). Vokey and Brooks (1992; Exp. 2) attempted to manipulate similarity and grammaticality independently. In their experiment, participants studied eight grammatical training strings composed of three to eight consonants. Half of the test strings (both grammatical and ungrammatical) were similar to a training string (called near items), and half were dissimilar to the training strings (called far items). The similarity manipulation involved changing specific letters to alter the similarity relation for pairs of strings. A near item, for example, was the same as a training item but with one letter changed. Far items had more letters changed, but—and this is the important part of the manipulation—both near and far items could be either grammatical or not. By constructing test strings in this way, Vokey and Brooks were able to construct a factorial arrangement of similarity and grammatical status. Their stimuli can be found in Vokey and Brooks' Table 4.

The experiment was administered in the usual fashion: After studying the training strings, the participants were told about the use of rules and, then, were instructed to decide whether or not novel test strings conformed to the rules. Half of the test items were grammatical, and half were ungrammatical.

Vokey and Brooks (1992) anticipated that performance would be dominated by their similarity factor. To their evident surprise, both the similarity and grammatical status variables influenced performance. Their results are summarized in the top panel of Figure 7.

The fact that both variables affected performance may appear to challenge the principles of

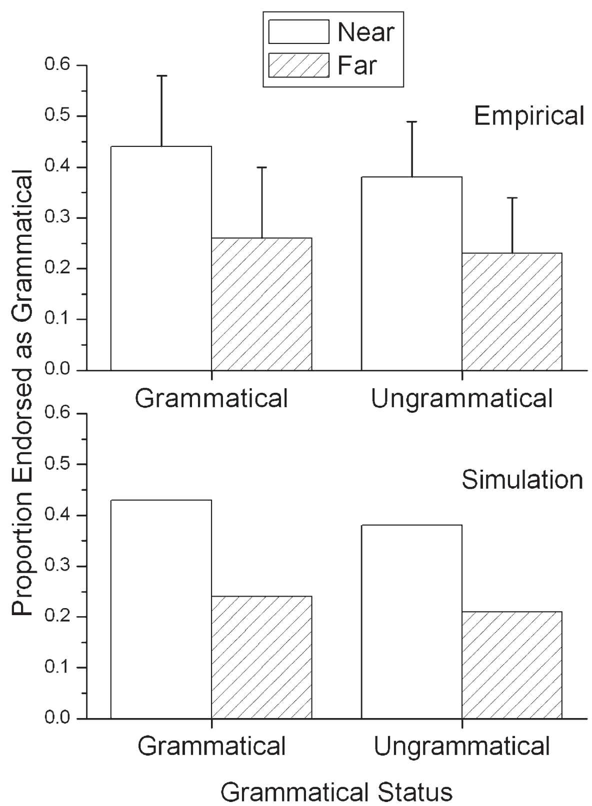

Figure 7. *Proportion of strings endorsed as grammatical as a function of grammatical status and similarity (near and far). The top panel summarizes data from Vokey and Brooks (1992). The bottom panel summarizes the simulation of the experiment. Parameters:* n = *20,* L = *.35,* k = *.371.*

the current model. The model's success in the earlier simulations implies that grammatical status is not an effective variable; the effective variable is global similarity. Global similarity differences occur because the grammar creates frequency differences among the stored exemplars, and the model's measure of global similarity is sensitive to the resulting similarity structure. Because global similarity is the effective variable underlying performance, Vokey and Brooks (1992) near–far manipulation cannot have made similarity entirely orthogonal to grammatical status.

The situation, here, is similar to that in the Dienes et al. (1991) example. The near–far manipulation was arranged by changing letters in specific pairs of stimuli. The model's decision, however, is based on a global measure of the match of the probe to memory. Based on the simulation of Dienes et al.'s materials, the global measure should be able to balance the local effects against the broader effect of the valid contingencies in memory.

To adapt the model to the Vokey and Brooks (1992) experiment, we constructed five random 20-element character vectors, one for each of the five letters: *M*, *R*, *V*, *X*, and *T*. The character vectors were concatenated to form a set of 72 vectors of 160 elements, one for each of the training and test strings used in the experiment. We stocked memory with the eight training strings, one row per string. Hence, after encoding the training strings, the memory matrix had eight rows (one for each training string) and 160 columns. As before, to acknowledge the difficulty of encoding the strings given fleeting exposure to them, we set L to a low value ($L = .35$).

A test string was endorsed as grammatical if the echo intensity exceeded the decision criterion—that is, $k = .371$. To obtain stable data, the results were averaged over 100 independent replications. The simulated results are presented in the bottom panel of Figure 7.

Comparing the two panels of Figure 7, it is clear that the simulation matched the data well. In particular, it captured the apparent although unreliable interaction between grammatical status and similarity, $RMSE = .016$. Decision in the model is based on global similarity, and, as in the Dienes et al. (1991) example, global similarity balanced the local perturbations against the contingencies among the stored exemplars. In short, the factorial arrangement of local similarity (near and far) and grammatical status is confounded with global similarity, and the model's reliance on global similarity allows it to capture the effect of grammaticality.

GENERAL DISCUSSION

After studying grammatical exemplars, participants can discriminate novel grammatical items from novel ungrammatical items. Because participants cannot articulate the grammar, the ability has been taken as evidence that they learn the grammar implicitly. We have reported three experiments that confirm standard results for illustrating sensitivity to structure. Experiment 1 used rules that imposed positional dependencies on symbols (i.e., the odd/even rule). Experiment 2 used rules that imposed sequential dependencies on symbols (i.e., a sequential grammar). Experiment 3 used sequential rules and varied both the amount of grammatical redundancy and the amount of learning. For all three experiments, the greater the number of rule violations in the probe, the more likely participants were to reject it.

We adapted an exemplar model of memory, Minerva 2, to help understand participants' performance in the artificial-grammar learning task. The model provided a close fit to performance in each of the three experiments and to three classic experiments conducted in other labs (Dienes et al., 1991; Reber, 1967; Vokey & Brooks, 1992). According to our account, participants store each studied stimulus in memory and judge the grammaticality of a novel stimulus in terms of its global similarity to the stored items. The model handled performance with materials constructed using positional and sequential dependencies without changing its representation, encoding, or retrieval assumptions. We conclude that evidence from standard artificial-grammar-learning tasks does not force an account based on a separate specialized implicit-learning system.

Instead, our evidence is that participants judge the grammatical status of a test probe by assessing its similarity to noisy representations of studied exemplars stored in memory. The larger implication of our claim is that performance in the judgement-of-grammaticality task is explained using the same principles and models as those used to explain performance in a sweep of other standard explicit-memory tasks.

Both the fragment and dual-systems positions postulate mechanisms to compile knowledge prospectively, either as rules or as a record of grammatical primitives (see Pothos, 2007, for a review of models applied to the judgement-of-grammaticality task). Instead, by focusing on global similarity in response to a probe, we have framed the judgement of grammatical status as a retrieval phenomenon. Retrieval is sensitive to structure of material in memory, even when individual exemplars are encoded imperfectly. We see no need to suppose that people abstract a distinct and separate representation of the grammar used to construct stimuli in our experiments.

Instead, we conclude that imperfect memory for training exemplars is enough to support the participants' ability to judge the grammatical status of test items. This empirical point, however, points to a broader message: Performance in the judgement-of-grammaticality task reflects the same principles widely used to understand performance in other memory tasks. As we documented earlier, Minerva 2 has been applied successfully to a wide number of situations. The present work ties performance in the artificial-grammar procedure to a long history of work in human memory. The integrative potential of such a simple model is particularly exciting.

Our general theoretical approach echoes the work of Pothos and Bailey (2000) who studied the role of similarity using unfamiliar visual stimuli. To assess similarity among the strings, they used a scaling technique recommended by Nosofsky (1991) and then applied those similarity measures to guide exemplar retrieval. We used a random vector to stand for each element in each stimulus string and calculated global similarity of each probe from the juxtaposition of events in memory: We did not attempt to represent the similarity relations among the elements of the string explicitly (e.g., *P* and *R* are more easily confused visually than *P* and *X*). Armed, however, with a scaling solution for similarity relations among symbols, it is certainly possible to construct vectors that map onto the similarity relations among the elements of each string. Our main point, however, is that performance reflects retrieval from memory and that similarity is constructed on the fly in response to a probe. One can go some distance without scaling subjective similarity ratings in advance.

A class of evidence for the separate systems view that we have not addressed concerns performance of brain-damaged patients. Reed, Squire, Patalano, Smith, and Jonides (1999) have showed that certain amnesiac patients perform well when discriminating grammatical from ungrammatical probes but cannot recall features from the studied material in an explicit cued-recall test. Because performance on the explicit task is impaired relative to that on the implicit task, the data are often cited as evidence for a division between explicit (called declarative) and implicit memory.

Two kinds of argument have been marshalled against the dissociation evidence. First, as we noted earlier, the dissociation does not have the logical force needed to compel a two-systems view (Dunn & Kirsner, 1988; Hintzman, 1990). More directly, Nosofsky and Zaki (1998; see also Zaki, Nosofsky, Jessup, & Unversagt, 2003) explained the dissociation between recognition and classification dissociation in amnesia using an exemplar model of categorization. Kinder and Shanks (2001, 2003) developed a related explanation using a connectionist model of memory to show that a generalized encoding deficit predicts the dissociation.

In terms of the current model, Reed et al.'s (1999) dissociation is not difficult to understand. In all of our experiments, encoding was difficult. To acknowledge the difficulty, we assumed that the encoded information was incomplete; each studied string in the model included a large number of indeterminate features: about 80% for Experiment 1, 45% for Experiment 2, and 55% for Experiment 3. As a result, the memory matrix was sparse. It is unlikely that any single string was represented well enough in the model to support retrieval for recall. As the simulations have shown, however, to assess the grammatical status of a test probe, one does not need strong (well-encoded) information for each studied item. Rather, the assessment of grammatical status depends on an aggregate constructed by summing across all the studied items. Because the aggregate sums information across the encoded representations, it supported classification even though the items in memory that the aggregate was formed from were sparse. In terms of the current work, the dissociation reflects the difference in retrieval required by the two tasks and should not be taken as evidence for a division between implicit and explicit learning.

The point has a larger implication. If deficits in retrograde amnesia reflect a generalized encoding deficit, rehabilitation techniques aimed at exploiting an intact—but implicit—memory system are misguided. Instead, rehabilitation should aim to help patients compensate for their deficits by training them to encode "strongly" and to make use of their residual capabilities to rely on their intuitions derived from retrieval of an aggregate. This suggestion requires development. However, it offers a novel perspective on therapeutic technique to help those with memory disorders.

A final point concerns the way decision is made in the artificial-grammar task. Standard theory is that participants endorse a probe if global similarity is large enough (e.g., Green & Swets, 1966). An alternative is that participants actively reject a probe that contradicts events in memory. The two alternatives are usually so highly correlated that they are hard to separate, but active rejection has been documented both in judgement of grammatical status (Wright & Burton, 1995) and in studies of recognition memory (e.g., Johns & Mewhort, 2002; Mewhort & Johns, 2000, 2005). Further research is needed to explore the role of contradiction to the study set and the nature of the decision process that people engage in when judging an item's grammatical status.

In summary, our account makes explicit a chain of relations. A grammar specifies rules for how grammatical strings are to be constructed. Exemplars produced using the rules are constrained, and, for that reason, the rules introduce frequency differences among the components of the grammatical strings. Because retrieval is sensitive to frequency differences among items in memory, people are able to judge the grammatical status of novel exemplars.

Original manuscript received 11 September 2007
Accepted revision received 23 February 2008
First published online day month year

REFERENCES

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language*, *39*, 371–391.

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

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

Bruner, J., Goodnow, J., & Austin, A. (1956). *A study of thinking*. New York: Wiley.

Clark, S. E. (1997). A familiarity-based account of confidence–accuracy inversions in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 232–238.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37–60.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

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

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.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*, 180–209.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*, 91–101.

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

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1–67.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

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

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

Hintzman, D. L. (1987). Recognition and recall in MINERVA-2: Analysis of the "recognition failure paradigm. In P. E. Morris (Ed.), *Modelling cognition* (pp. 215–229). London: Wiley.

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

Hintzman, D. L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology*, *41*, 109–139.

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.

Johns, E. E., & Mewhort, D. J. K. (2002). What information underlies correct rejections in short-term recognition memory? *Memory & Cognition*, *30*, 46–59.

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

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

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

Kinder, A., & Shanks, D. R. (2003). Neuropsychological dissociations between priming and recognition: A single-system connectionist account. *Psychological Review*, *110*, 728–744.

Knowlton, B. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, *3*, 172–179.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 79–91.

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.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of non-local musical rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1417–1432.

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review*, *12*, 703–710.

Kwantes, P. J., & Mewhort, D. J. K. (1999). Modeling lexical decision and word naming as a retrieval process. *Canadian Journal of Experimental Psychology*, *53*, 306–315.

Kwantes, P. J., & Neal, A. (2006). Why people underestimate y when extrapolating in linear functions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1019–1030.

Manza, L., & Reber, A. S. (1997). Representing artificial grammars: Transfer across stimulus forms and modalities. In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 73–106). Oxford, UK: Oxford University Press.

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.

McAndrews, M. P., & Moscovitch, M. (1985). Rule-based and exemplar classification in artificial grammar learning. *Memory & Cognition*, *13*, 469–475.

Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. *Psychological Review*, *85*, 207–238.

Mewhort, D. J. K., & Johns, E. E. (2000). The extralist-feature effect: Evidence against item matching in short-term recognition memory. *Journal of Experimental Psychology: General*, *129*, 262–284.

Mewhort, D. J. K., & Johns, E. E. (2005). Sharpening the echo: An iterative resonance model for short-term recognition memory. *Memory*, *13*, 300–307.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology*, *56*, 433–491.

Murdock, B. B. (1982a). Recognition memory. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 1–26). New York: Academic Press.

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

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

Neal, A., & Hesketh, B. (1997). Episodic knowledge and implicit learning. *Psychonomic Bulletin and Review*, *4*, 24–37.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorisation relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 3–27.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorisation and recognition in amnesiacs and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*, 247–255.

Perruchet, P., & Gallego, J. (1997). A subjective unit formation account of implicit learning. In D. C. Berry (Ed.), *How implicit is implicit learning?* (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.

Perruchet, P., & Pacton, S. (2005). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Science*, *10*, 233–238.

Perruchet, P., & Vintner, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences*, *25*, 297–388.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review*, *91*, 281–293.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, *133*, 227–244.

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.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

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

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, *81*, 115–119.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.

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

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, *125*, 123–138.

Reed, J. M., Squire, L. R., Patalano, A. L., Smith, E. E., & Jonides, J. (1999). Learning about categories that are defined by object-like stimuli despite impaired declarative memory. *Behavioral Neuroscience*, *113*, 411–419.

Rossnagel, C. S. (2001). Revealing hidden covariation detection: Evidence for implicit abstraction at study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1276–1288.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*, 172–196.
Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 592–608.
Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.
Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
St. John, M. F., & Shanks, D. R. (1997). Implicit learning from an information-processing standpoint. In D. C. Berry (Ed.), *How implicit is implicit learning?* (pp. 162–194). Oxford, UK: Oxford University Press.
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.
Vokey, J. R., & Brooks, L. R. (1994). Fragmentary knowledge and the processing-specific control of structural sensitivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1504–1510.
Vokey, J. R., & Higham, P. (2005). Abstract analogies and positive transfer in artificial grammar learning. *Canadian Journal of Experimental Psychology*, *59*, 54–61.
Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, *122*, 227–248.
Whittlesea, B. W. A., & Wright, R. L. (1997). Implicit (and explicit) learning: Acting adaptively without knowing the consequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 181–200.
Wright, R. L., & Burton, A. M. (1995). Implicit learning of an invariant: Just say no. *Quarterly Journal of Experimental Psychology*, *48A*, 783–796.
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.
Zaki, S. R., & Nosofsky, R. M. (2001). A single-system interpretation of dissociations between recognition and categorisation in a task involving object-like stimuli. *Cognitive, Affective, and Behavioral Neuroscience*, *1*, 344–359.
Zaki, S. R., Nosofsky, R. M., Jessup, N. M., & Unversagt, F. E. (2003). Categorisation and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society*, *9*, 394–406.

APPENDIX

Comparison to simulations by Dienes (1992)

We adapted Minerva 2 to the artificial-grammar task to model Dienes et al.'s (1991) experiment; the simulations worked well. Dienes (1992) adapted Minerva 2 to Reber's (1967) grammar-learning experiment with mixed results. Why did our adaptation work when his did not? We think the difference reflects the way in which stimuli are represented in the two applications of the model.

To code our stimuli, we first constructed a random vector to represent each character and then concatenated the appropriate character vectors to form a string. Thus, a character's identity was indicated by a pattern of $+1/-1$ elements within a character field, and the character's position within the string was indicated by its position in the series of concatenated character vectors. Imperfect encoding was accomplished by replacing some of the vector elements with zeros to indicate that the element was unknown. Replacing elements in this way obscures information about each item independently without impacting memory for each symbol's position in the string.

Dienes' (1992) coding system was very different. Reber's (1967) grammar involved strings of letters (up to six) taken from a population of five letters. In Dienes' simulation, the letters were represented by a set of five input units, one unit for each letter in the population. Each letter was coded by setting one unit within the set to a value of $+1$ and setting the other units to values of -1. Using this system, the letters M and V might be coded $\{+1\ -1\ -1\ -1\ -1\}$ and $\{-1\ +1\ -1\ -1\ -1\}$, respectively. A string of letters was coded by setting the identity of sets of units that were hardwired to a position. That is, the identity of each letter was coded within a set of five units, and a letter's position in the string was coded using sets of five units, one set of five for each position in a string. In terms of the example, if the first set of five units indicates the letter in Position 1, and the second set of five units indicates the letter in Position 2, MV would be coded: Position 1 = { +1 −1 −1 −1 −1}, Position 2 = {−1 +1 −1 −1 −1}.

In light of the fleeting exposure to each study string, participants were unlikely to encode them very well; indeed, this was the justification for introducing so much noise when encoding materials in our simulations. Suppose that a participant could not remember what letter was in a particular position in a string. To represent the situation, the model must be able to degrade the identity of the letter without introducing other complications. In our adaptation of Minerva 2, such distortion was easy to implement by replacing some of the +1/−1 values with 0s. In Dienes' (1992) adaptation, however, the distortion is not easy to implement. Consider the { +1 −1 −1 −1 −1} vector coding for M in our example. Replacing the +1 value by a 0—that is, {0 −1 −1 −1 −1}—may appear to degrade the identity of the letter, but, because the −1 values indicate what the letter is not, the vector still signals M. Replacing one of the other −1 values to the +1 value introduces competition between two possible letters, but does not degrade the representation of either one. For example, {0 0 −1 −1 −1} indicates the letter is either M or V (because it is not one of the other three possibilities). Unlike our implementation, replacing elements in this way does not obscure information about each item independently.

The quality of encoding is important: Our simulations acknowledge that participants in the artificial-grammar task have only a fleeting opportunity to encode the stimulus exemplars. If we were to ignore that fact and simulate performance assuming perfect encoding, the simulations fail, as they should. In Experiment 1 with $L = 1.0$ (rather than $L = .2$), for example, an ungrammatical string will be judged to be grammatical only rarely (if at all), a result inconsistent with the data. Although the effect of perfect encoding is most dramatic for our Experiment 1, the same is true for our other two experiments. The influence of quality of encoding has on performance is well established in work on human memory (e.g., the shorter exposure to a stimulus is, the worse memory for it becomes).

In fairness, Dienes' (1992) goal was not to show Minerva 2 in the best possible light but to contrast it with several other accounts, each implemented in a neural network. To make the comparisons, all of the models were developed using the same representation assumptions. Hintzman (1990) had sketched how Minerva 2 could be adapted to the neural-network formalism, and Dienes borrowed heavily from Hintzman's sketch. That said, to make Minerva 2 compatible with the other models, Dienes introduced constraints that our implementation escapes.