Does stimulus appearance affect learning?

EMMANUEL M. POTHOS
University of Crete

NICK CHATER
University of Warwick

ELENI ZIORI
University of Ioannina

We examined the learning process with 3 sets of stimuli that have identical symbolic structure but differ in appearance (meaningless letter strings, arrangements of geometric shapes, and sequences of cities). One hypothesis is that the learning process aims to encode symbolic regularity in the same way, largely regardless of appearance. Another is that different types of stimuli bias the learning process to operate in different ways. Using the experimental paradigm of artificial grammar learning, we provided a preliminary test of these hypotheses. In Experiments 1 and 2 we measured performance in terms of grammaticality and found no difference across the 3 sets of stimuli. In Experiment 3 we analyzed performance in terms of both grammaticality and chunk strength. Again we found no differences in performance. Our tentative conclusion is that the learning process aims to encode symbolic regularity independent of stimulus appearance.

It is a trivial statement to point out that the aim of the learning process is to identify regularity in the stimuli that are learned. It is equally trivial to observe that there are many kinds of regularity. For example, in language learning it is often assumed that the cognitive system is able to infer a rulelike representation of the grammatical and syntactical structure in linguistic input (e.g., Chomsky, 1965; Elman, 1996; Marcus, 2001). It makes sense for the cognitive system to seek rulelike regularities in language because otherwise the features that make language useful (e.g., compositionality, productivity, systematicity; Fodor & Pylyshyn, 1988; for an early proposal see Boole, 1854) could not be supported; even neural network models of language have been shown to carry out linguistic operations on the basis of structures that look like rules (Dienes, 1992; Pothos, 2005b). The fact that linguistic input (at least when it comes to written text; cf. Saffran, Aslin, & Newport, 1996) has a salient symbolic structure presum-

AMERICAN JOURNAL OF PSYCHOLOGY
Summer 2006, Vol. 119, No. 2, pp. 277–301
© 2006 by the Board of Trustees of the University of Illinois

ably helps the cognitive system to search for corresponding grammatical or syntactical rules.

By contrast, many of our everyday categories do not have an obvious representation in terms of rules. Rosch and Mervis (1975) argued that some stimuli are recognized as members of a category if they share more attributes than other stimuli (see also Pothos & Chater, 2002). Several other researchers have proposed and successfully defended approaches to categorization whereby the classification of a novel instance of a category is determined either by similarity to the category's exemplars (Kruschke, 1992; Nosofsky, 1989) or by similarity to the category's prototype (Hampton, 1999; Homa & Chambliss, 1975). This is not to say that there have not been attempts to understand certain types of categories in terms of rules (e.g., Malt, 1994; Pothos & Hahn, 2000; Rips, 1989, 2001) but rather that we can identify several situations whereby the learning process leads to similarity knowledge (more accurately, to knowledge that allows judgments of similarity between a novel instance and experience).

In short, there are at least two distinct outcomes of the learning process: knowledge in the form of rules and knowledge in the form of similarity (Pothos, 2005b). Moreover, certain types of stimuli appear to be served better by knowledge of rules, others by similarity. We can now frame the hypotheses we want to examine in the present work. On one hand, it is possible that the learning process simply tries its best to encode the regularity in a set of stimuli. If this regularity is best represented in terms of rules, then it is rules that the learning process will abstract. If it is best represented by similarity comparisons, then the learning process will encode the knowledge that allows such comparisons (e.g., whole instances, prototypes). On the other hand, it is possible that different types of stimuli bias the learning process to anticipate different kinds of regularity. For example, when faced with stimuli that have a very salient symbolic structure, the learning process may be biased to look specifically for rules. It is the purpose of this work to provide a preliminary investigation of this issue.

We chose to use the experimental paradigm of artificial grammar learning (AGL; Chomsky & Miller, 1958; Reber, 1967). AGL concerns the learning of symbolic sequences generated from a finite state language. A finite state language for a given AGL experiment is a set of rules that determine the allowed continuation relationships in the symbols of the stimuli in the experiment (Figure 1). Using AGL is advantageous for two reasons. First, AGL performance can be characterized in a way that allows different inferences about the knowledge that is learned. Second, we have a very long tradition of similar studies, so that methodological issues and possible confounds are reasonably well understood.

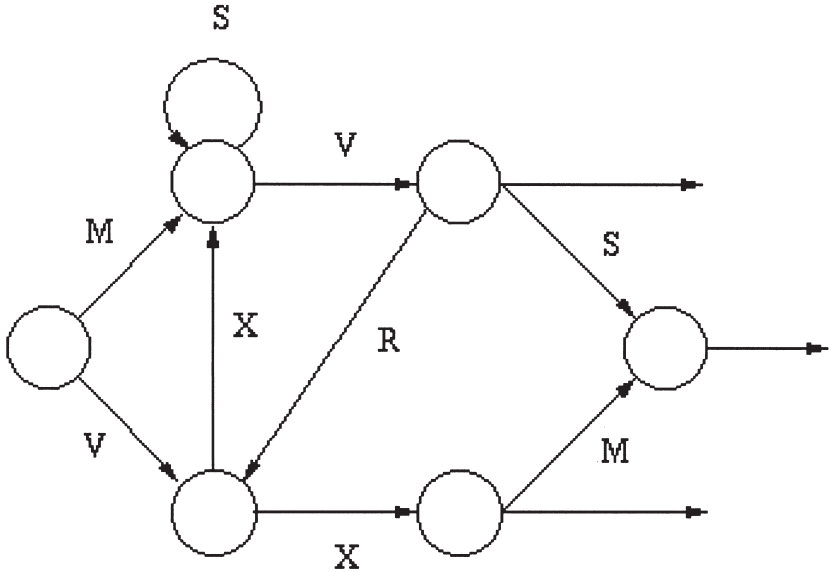

Figure 1. The finite state language used in Experiments 1 and 2. Each of the allowed transitions is associated with the addition of a symbol, so in going from the entry state to any of the exit states on the diagram, different sequences of symbols can be constructed. For example, *M-S-V* would be a grammatical sequence, but *V-V-X* would not be because once a *V* has been added to a sequence, there is no transition that could lead to the second *V*. In this example, the symbols associated with each allowed transition in the finite state grammar are letters, but there is no restriction as to what they could be

Sequences of symbols that are consistent and not consistent with the rules of a finite state language are called grammatical (G) and not grammatical (NG), respectively. In a typical AGL experiment, the experimenter first identifies all the G sequences that can be generated by a finite state language; these are typically between 40 and 60 (if repetition loops are restricted in some appropriate way). Participants are first shown a subset of the G sequences with the instructions to simply observe them. In the test phase, participants are told that all training sequences were generated by a complex set of rules. They are then presented with the remaining G sequences, together with NG ones, and they are asked to identify which are which (identifying a test sequence as G or NG is called a grammaticality decision). Reber (1967; see also Miller, 1958) first used this paradigm to find that participants could indeed discriminate between G and NG sequences in test with above-chance accuracy, a finding that was extensively

replicated in later research (Brooks & Vokey, 1991; Dulany et al., 1984; Gomez & Schvaneveldt, 1994; Higham, 1997; Higham, Vokey, & Pritchard, 2000; Knowlton & Squire, 1994; Mathews et al., 1989; Meulemans & van der Linden, 1997; Mathews et al., 1989; Perruchet & Pacteau, 1990; Pothos & Bailey, 2000; Shanks, Johnstone, & Staggs, 1997; Whittlesea & Dorken, 1993; for discussions of AGL results see Dienes, 1992; Reber, 1989; Redington & Chater, 1996).

A crucial aspect of AGL methods is that the format of the symbols that make up the sequences in an AGL experiment can be arbitrary. In Figure 1, the symbols are letters, but they could also be graphic symbols, colors, or musical tones (Altmann, Dienes, & Goode, 1995; Chan, 1992; Pothos & Bailey, 2000). By creating stimuli that differ in appearance but share an identical underlying structure, we can examine whether the learning process is influenced by appearance or is simply guided by the underlying structure. It is unlikely that AGL performance is entirely independent of stimulus appearance. The issue we believe that we can validly pursue is to examine whether the effect of stimulus appearance is important enough to lead to statistically significant changes in participants' performance, in the context of standard aspects of AGL methods.

AGL theories

In this section we review existing AGL theories with a view to arguing for the suitability of AGL as a paradigm for the current investigation. Reber's view (1967, 1989) was that because the G versus NG distinction is defined by rules and because participants can successfully discriminate (to some extent) between G and NG sequences, participants must have acquired some knowledge of the rules that underlie the G sequences. Dulany et al. (1984; Dulany, Carlson, & Dewey, 1985) have developed Reber's rule hypothesis of AGL performance by providing an essentially introspective method to identify what kind of rules each participant used in his or her grammaticality decisions. These investigators were able to show that the validity of the rules reported by each participant correlated highly with his or her grammaticality accuracy.

A number of AGL investigators have pointed out that other kinds of knowledge might be the basis for participants' grammaticality decisions in an AGL experiment. For example, Vokey and Brooks (1992) and Higham et al. (2000) observed in their experiments that participants were biased to select as G the test sequences that were more similar to the training ones as well as the test sequences that were actually G. Similarity between two sequences was assessed in terms of edit distance, that is, the number of symbol changes needed before the two sequences are identical (cf. Chater & Hahn, 1997). Likewise, Pothos and Bailey (2000) reported separate influences of grammaticality and similarity on AGL performance; how-

ever, in that study similarity was modeled on the basis of actual similarity judgments provided for all the stimuli in Pothos and Bailey's experiment. Perruchet and Pacteau (1990; Perruchet, Vinter, Pacteau, & Gallego, 2002; see also Knowlton & Squire, 1996) suggested that the test items that are more likely to be selected as G are the ones that contain fragments (i.e., pairs or triplets of letters) that are familiar from training. Finally, Dienes (1992; Dienes, Altmann, & Gao, 1999) successfully provided neural network simulations of AGL performance. In sum, there appears to be a consensus that AGL performance is guided by several kinds of knowledge, such as rules, similarity, and associative knowledge (Johnstone & Shanks, 1999; Pothos & Bailey, 2000; for a discussion see Pothos, 2005b).

Theoretical effort in AGL thus clearly emphasizes the importance of regularity at the symbolic level, with no reference to potential influences from stimulus appearance. Even Vokey and Brooks's (1992) similarity approach involves a similarity model that is specified in terms of the symbolic structure of the AGL sequences. Of course, Pothos and Bailey (2000) modeled AGL performance in terms of similarity judgments, which might very plausibly depend on the format of the stimuli. But Pothos and Bailey were more interested in empirically establishing the relevance of similarity in AGL than in providing a specific account of how similarity influences could be modeled—so their work is uninformative as to whether stimulus appearance affects the learning process.

Overall, therefore, we can see that in AGL there is a potential for different kinds of stimuli to bias the learning process to look for different kinds of regularity. However, this possibility is not taken into account in existing AGL theory. Accordingly, the present investigation is primarily an exploratory one: Although we shall try to motivate our manipulations, our intention is to create different kinds of stimuli and see whether they affect performance.

Previous work

Cleeremans and McClelland (1991) provided an ambitious demonstration that the learning process can identify regularity in stimuli generated from a finite state language when the stimuli are not presented as letter strings. Their participants completed a serial reaction task in which the same target stimulus appeared in different locations on the screen. It was the succession of appearance locations that was determined by finite state language dependencies or randomly. Cleeremans and McClelland's results clearly showed that participants were able to gradually respond faster in the first case.

Using the standard AGL paradigm, Altmann et al. (1995) showed that exposure to AGL stimuli instantiated in one format would lead to sensitivity to grammaticality distinctions in a test part, where the stimuli were

instantiated not only in a different format but also in a different modality. For example, they found that exposure to training stimuli as musical tones led to grammaticality sensitivity with test stimuli as sequences of letters (see also Chan, 1992, for similar results with graphic stimuli and Whittlesea & Wright, 1997, for a more recent similar investigation). However, because Altmann et al. did not compare performance with different kinds of stimuli, it is not possible to conclude from their results whether different kinds of stimuli might influence the learning process.

Pothos and Bailey (2000) did compare grammaticality accuracy with three different types of stimuli: sequences of geometric shapes, geometric shapes arranged so that shapes corresponding to later symbols in a sequence enclosed all previous ones, and lines arranged so that different symbols corresponded to the angles between the individual straight lines that made up a stimulus (e.g., if *V* corresponded to 75° and *X* to 150°, to create the string *VX* we would need a line segment at 75° relative to the horizontal and another line segment at 150° relative to the first one, or actually 45° to the horizontal). Overall, grammaticality accuracy was compared on the basis of a one-way ANOVA, and no overall differences were identified. However, Pothos and Bailey did not carry out a more detailed comparison of performance with the three kinds of stimuli they considered.

There are at least two studies examining ways in which the learning process may be biased by explicit methodological manipulations. In their Experiment 1 Reber and Perruchet (2003) compared performance of two sets of stimuli having identical symbolic structure but differing in terms of whether the stimuli were created using high-frequency French consonants or low-frequency ones. For the group of participants who went through a training phase, grammaticality accuracy was nearly identical regardless of whether the low- or high-frequency consonants were used. Pothos (2005b) manipulated the appearance of the stimuli and the experimental instructions to induce specific prior expectations about the structure of the stimuli. For example, in one experiment the stimuli appeared as sequences of cities (a manipulation presently used as well) that corresponded to the routes of a traveling salesman. Participants were told that the salesman had to plan his routes so that he traveled to cities as near as possible, or his journeys would be inefficient. In one condition of the experiment, the G sequences involved cities closer together, and in another they did not (so that the structure of the G stimuli was contrary to participants' expectation). Pothos found that participants' performance was impaired in the latter condition (see also Perruchet, Pacteau, & Gallego, 1997, for how prior knowledge may influence sensitivity to covariation; cf. Kinder, Shanks, Cock, & Tunney, 2003).

Although such research is informative in regard to the flexibility of the

learning process, at present we are interested in whether the learning process would be spontaneously (that is, regardless of prior knowledge or explicit considerations) biased to look for specific kinds of stimulus structure depending on how stimuli look.

Methodological issues

In Experiments 1 and 2 we chose to replicate as closely as possibly Reber and Allen's (1978) experiment. The training and test stimulus sets created by these investigators are used often in AGL studies. Therefore, in choosing to base Experiments 1 and 2 on Reber and Allen's material, we aimed for an investigation that would be as comparable to previous research as possible. A particular advantage of Reber and Allen's materials is that the NG test items were created by systematically introducing violations to G sequences. Thus it is possible to analyze performance with these materials both in terms of overall grammaticality accuracy and in terms of grammaticality accuracy in each of the test item subsets. In the present investigation, we examined participant performance in terms of grammaticality accuracy because grammaticality accuracy is the most widely used measure of AGL performance. Arguably, there might be a case for using other measures of performance (e.g., from detection theory), but the advantage of using grammaticality accuracy is that it enhances comparability with previous research.

Reber and Allen (1978) viewed grammaticality accuracy as a measure of rule learning. Since their investigation, however, there have been several demonstrations that grammaticality accuracy with their materials may covary with other kinds of knowledge, such as knowledge of string fragments (Perruchet & Pacteau, 1990) or similarity (Vokey & Brooks, 1992). Thus, in Experiments 1 and 2 we cannot infer whether stimulus format biases the learning process to look, for example, for rules or similarity (grammaticality accuracy with the Reber and Allen materials is best seen as an aggregate measure of different sources of knowledge). Accordingly, in Experiments 1 and 2 our only purpose was to examine whether stimulus appearance may sometimes inhibit the operation of the learning process.

This shortcoming of Experiments 1 and 2 is addressed in Experiment 3, in which we used the materials of Knowlton and Squire (1996, Experiment 1). Knowlton and Squire created the test sequences so that the G ones were balanced in terms of global associative chunk strength relative to the NG ones. A chunk is a bigram or a trigram. To compute the global associative chunk strength of a test item, first divide it into all possible chunks. Then compute the associative strength of each chunk, which is its frequency of occurrence in the training items. The average associative chunk strength of all chunks in a test item is the global associative chunk

strength of the item. Knowlton and Squire motivated their measure of chunk strength from Servan-Schreiber and Anderson's (1990) competitive chunking hypothesis, which is effectively a hypothesis about learning based on associative learning principles. However, global associative chunk strength has been interpreted as a measure of similarity both by Knowlton and Squire and in later discussions (Pothos, 2005b). In Knowlton and Squire's materials, because of the care with which possible similarity effects were controlled for, grammaticality can be more straightforwardly understood as reflecting rule knowledge (see also Pothos & Bailey, 2000). By using Knowlton and Squire's materials, we can more directly examine whether stimulus appearance biases the learning mechanism to look for rule knowledge or similarity knowledge.

A controversial issue in AGL is how to establish whether learning really takes place. In early studies, grammaticality accuracy was compared against chance performance, the rationale being that unless participants learned something in the training phase they would be unable to distinguish correctly between G and NG sequences. However, in some cases investigators have reported that no-training control participants could distinguish G sequences from NG ones with above-chance accuracy (for a review of relevant issues see Redington & Chater, 1996). It has been suggested that the only way to establish learning in an AGL experiment is to compare the performance of no-training control participants with that of experimental participants who had the benefit of a training phase. Such an approach has also been shown to be inadequate because even when no training controls perform at the same level as experimental participants, the basis for the G endorsements of the experimental participants is different from that of the control ones (Reber & Perruchet, 2003; for a different view see Dienes & Altmann, 2003, and for further discussion of this issue see Pothos, 2005a; Pothos & Bailey, 2000). This complex issue is beyond the scope of this investigation, and we have not tried to directly establish whether learning took place in our experiments. Instead, we used materials with which learning has been demonstrated in the past. Note that AGL research so far suggests that it is unlikely that the learning phase has no influence on the test phase (Pothos & Bailey, 2000), with the possible exception of stimulus sets that have particular confounds (Redington & Chater, 1996).

In this work we chose to study three kinds of stimuli: standard letter strings like the ones used by Reber and Allen (1978) and in the majority of AGL studies; sequences of cities (city stimuli) that corresponded to the routes of an airline, in which each finite state language symbol corresponded to a city (cf. Pothos, 2005a); and embedded sequences of geometric shapes (shape stimuli), in which the first shape symbol appeared in the center of the stimulus arrangement, the next one enclosed the first,

and so forth, so that the last shape symbol would enclose all others (see Figure 2 for examples).

With the city stimuli we hoped to examine possible influences of background knowledge on AGL performance. For example, starting an airplane route in London might be illegal according to the finite state language, but it is certainly not for the associations participants bring into the laboratory (cf. Pothos & Cox, 2002, for an analogous AGL investigation in a clinical setting). Would background expectations and biases about which routes "make sense" affect participants' grammaticality judgments? If that is the case, then grammaticality performance with the city stimuli would be at chance because these stimuli were not constructed to reflect any realistic constraint about which airplane routes might be more plausible than others. Note that this issue is different from that examined by Pothos (2005a) because in that examination there was a direct manipulation of the actual structure of the stimuli and participants' a priori expectations (which were instructionally induced) for the stimuli.

With the shape stimuli our aim was to create stimuli that could be perceived as individual objects (as suggested by the Gestalt perceptual principle of closure) rather than collections of shapes. Note that sequential information between the constituent elements of stimuli is as accessible (in principle) to participants as sequential information with, for example, the letter string stimuli. With the letter stimuli, participants may observe, for example, that letter *S* often follows letter *M*. With the shape stimuli, equally, participants might observe that a square often encloses a circle. However, would the possible impression of the shape stimuli as individual objects bias the learning process to emphasize learning in terms of similarity as opposed to learning in terms of rules?

This section has provided some simple conceptions of how stimulus appearance might affect the learning process. Of course, it is possible

MSSSV

Athens --> London --> London --> London --> Berlin

Figure 2. Examples of the types of stimuli used

that the learning process simply aims to encode stimulus regularity in the most efficient way regardless of stimulus appearance.

EXPERIMENT 1

METHOD

Design

This experiment consisted of three AGL tasks that differed only with respect to the format of the stimuli. In one task the stimuli appeared as letter strings (letter condition), in another as sequences of cities that corresponded to the routes used by an airline company (city condition), and in the final task as embedded geometric shapes that were meant to give the impression of whole objects (shape condition). Reber and Allen (1978) reported results of adequate power with 10 participants, and likewise we chose to test 10 participants in each task. Participants were all University of Oxford students who took part for a payment of £**2** each.

Materials

The letter string AGL task was identical to that of Reber and Allen (1978); Figure 1 shows Reber and Allen's finite state language, one of the most widely used finite state languages in AGL (e.g., Dulany et al., 1984; Perruchet & Pacteau, 1990). In brief, there are 20 training items, 25 G test items, and 25 NG test items. To create the city stimuli we mapped the letters used for the letter stimuli to cities; *R, X, M, S,* and *V* were mapped to Athens, Berlin, London, Paris, and Madrid. For the shape stimuli, *R, X, M, S,* and *V* were mapped to a circle, a hexagon, a concave shape, a square, and a rhombus. The shapes were chosen so that they were of roughly similar salience; likewise, the cities were intended to be well-known European capitals and representative of most of Europe. Examples of the stimuli used are shown in Figure 2.

All stimuli were presented individually on A4–sized paper and were printed in black ink. The size of the stimuli ranged from approximately 3 cm × 1 cm to 6.5 cm × 1 cm for the letter strings, 3 cm × 4 cm to 15 cm × 15 cm for the shapes, and 9.5 cm × 1 cm to 18.5 cm × 1 cm for cities. The shape stimuli were created so that the longest axis of the first shape was about 2 cm, and subsequent shapes were added so that they contained all previous shapes; for two consecutive figures (in the same stimulus), the distance between them was between 0.5 cm and 1 cm.

Procedure

We closely replicated the procedure of Reber and Allen (1978). In the first phase of the experiment, participants were presented with a folder containing the training stimuli and instructions. The instructions emphasized that participants had to "pay the utmost attention to the letter strings" (or cities or shapes, depending on the particular set of stimuli used) but nothing more; also, although participants were told that there would be a second phase, no information was given as to what would be required of them. In the case of the city stimuli, participants were told

that "city names correspond to where a plane of the company was at noon, on successive days." The training items consisted of 20 unique G items presented three times in a random order (two different orders were used). Once the instructions had been read, participants turned to the first sheet in the folder, which they saw for approximately 10 s. The experimenter tapped on the table to indicate that the participant had to proceed to the next item. On average, the training phase of the experiment took about 10 min.

Participants then received another folder and more printed instructions. The folder contained 50 unique items presented twice in a random order; half were G, and the other half violated the rules of the Figure 1 finite state language in specific ways. It was emphasized that the order of cities (or letters or shapes) in each of the training items was determined by a complex set of rules, that half of the items in the second folder complied with these rules whereas the other half violated them, and that participants were to distinguish between the two sets. Also, they were told that they must not look back at previous items when making their grammaticality decisions. In the case of the city stimuli, it was added that the rules were devised to "ensure that the available airplanes are put to best use" and that the routes were independent of each other. The "independence" remark was made to discourage participants from trying to guess whether the routes collectively might make sense in terms of some efficient travel scheme (cf. Pothos 2005a). Such an effort would be futile, of course, because the cities were randomly mapped to letters and might unfairly prejudice performance in the city stimuli. Participants indicated their grammaticality decisions by attaching a self-stick note with the letter *G* (*good*) or *N* (*not good*) to the corresponding items. There was no time limit in making any one decision, but participants were not allowed to make corrections on previous items.

RESULTS

Given the small sample size, the presence of even a single outlier could distort performance, so we decided to eliminate participants whose performance deviated by 2 standard deviations or more from the overall mean (one participant was eliminated in this way from the city condition). In Table 1 it can be seen that the grammaticality accuracy is nearly identical across stimuli. Additionally, we computed single-sample *t* tests (two tailed) against chance as an indication that grammaticality learning took place; all *t* tests were highly significant.

Table 1. Mean grammaticality accuracy in the three conditions of Experiment 1

Stimuli	Mean	*SD*	*t*-Test against chance
Letters	67.2%	7.6%	$t(9) = 6.9$, $p < .0005$
Shapes	63.6%	6.7%	$t(9) = 6.4$, $p < .0005$
Cities	64.8%	4.7%	$t(8) = 9.4$, $p < .0005$

A simple test of whether stimulus appearance affects AGL performance would be to compare grammaticality accuracy between the three conditions. The independent sample one-way ANOVA was not significant, $F(2, 26) = 0.76$, $p = .475$.

As mentioned earlier, Reber and Allen (1978) constructed their NG stimuli by violating G ones in specific ways. Therefore, the NG items could be divided into the following mutually exclusive subsets: items with violations in the first position, violations in the second position, deep violations, violations in the next to last position, violations in the last position, and items that were spelled backwards (there were roughly the same number of items in each of these groups). Also, five of the test G items were repeated from training. We might acquire some insight about whether participants were doing something different in each condition by examining their performance in each of these subsets of the test stimuli. Figure 3 shows these results. We can see that patterns of error for the NG items

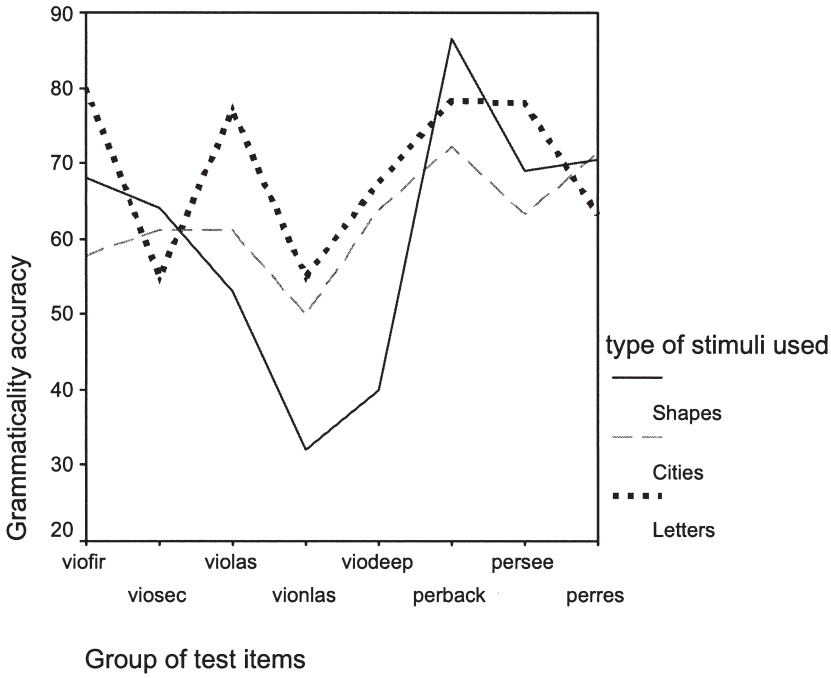

Figure 3. Grammaticality accuracy in different subsets of the test items in Experiment 1. viofir = violations in the first position; viosec = violations in the second position; violas = violations in the last position; vionlas = violations in the next to last position; viodeep = deep violations; perback = items spelled backwards; persee = G test items repeated from training; perres = novel G test items

are nearly identical for the city and letter stimuli. However, in the shape condition, participants were particularly bad with items that had deep violations (i.e., violations in the middle of the sequences) and violations in the next to last position. It is also interesting to note that grammaticality accuracy was nearly identical in the three conditions with both the test G items repeated from training and the novel test G items. A mixed factorial design was run with stimulus type as a between-participants factor and performance for different subsets of test items (as shown in Figure 3) as a within-participants factor (SPSS Greenhouse–Geisser correction applied). The interaction was found not to be significant, $F(7.34, 95.5) = 1.71$, $p = .11$. Note that this ANOVA is slightly contrived because there was large variation in the number of items in each group (the group "rest of G items" had a lot more items than the other groups).

DISCUSSION

Our results indicate that stimulus appearance does not affect the learning process. Overall, grammaticality accuracy was nearly identical in all three cases. The grammaticality patterns of errors were very similar for the letter and city stimuli, suggesting that whatever biases participants might have had about which airline routes were more plausible, these biases did not influence AGL performance. With the shape stimuli it might appear that stimulus format enhanced perception of the initial and last symbols in each stimulus, at the expense of the intermediate symbols; however, this perception could not be statistically supported. Finally, despite the small sample sizes, performance in all three cases was significantly above chance.

A variation of Experiment 1 that might show an influence of stimulus appearance on the learning process involves the AGL transfer paradigm. In the standard AGL paradigm, both the training and test sequences are created from the same set of symbols. In transfer experiments, this is not the case so that, superficially, the test sequences look different from the training ones; for example, the training sequences might be made from *T, V, J,* and *X* and the test ones from *A, C, W,* and *K*. It has been widely demonstrated that participants can successfully discriminate between G and NG strings in such transfer experiments (e.g., Mathews et al., 1989; Vokey & Brooks, 1992; Whittlesea & Dorken, 1993; for an overview see Reber, 1989). There have been many accounts of transfer performance. Of relevance here is the observation that transfer performance requires participants to encode the training stimuli either in a way that goes beyond their superficial characteristics (e.g., in terms of their abstract symbolic structure; Reber, 1989) or in way that allows information about the abstract structure of the stimuli to be used in test (e.g., Vokey & Brooks, 1992).

It is possible that certain types of stimuli encourage such abstract representations and others do not (cf. Perruchet & Vinter, 2002). For example, with the letter stimuli there is evidence of extensive abstraction (in either of the two ways just mentioned). This is intuitive because letter stimuli not only have a highly salient symbolic structure but also are entirely meaningless and unfamiliar (presumably encouraging participants to see the logic behind the stimuli). By contrast, the shape stimuli are obviously perceptual in nature. This might bias the learning process to assign more importance to their exact form so that corresponding transfer performance would likewise be impaired. Moreover, in a transfer experiment with the city stimuli it is possible that general knowledge expectations might have a greater influence on performance because participants would not be able to directly use their experience of the training stimuli to process the test ones.

In Experiment 2 we attempted to investigate these possibilities using an AGL transfer paradigm.

EXPERIMENT 2

METHOD

Design

The design was identical to that of Experiment 1. However, because of the greater variability typically observed in transfer AGL experiments, 20 participants were tested with each type of stimulus. Participants were University of Oxford students and received £2 for taking part.

Materials

We used the Reber and Allen (1978) grammar, as before. However, in each of the three conditions, the symbols making up the stimuli were changed from training to test. In the letter case, the letters *R, X, M, V,* and *S* were used in the test phase and *F, Q, J, C,* and *D* in the training phase. In the city case, Athens, Berlin, London, Paris, and Madrid were used in the test phase and Boston, New York, Atlanta, Miami, and Chicago in the training phase. In the shape case, a circle, a hexagon, a concave shape, a square, and a rhombus were used in the test phase and a square with ears, a square with triangles, a dodecagon, and ellipses arranged like a cross were used in the training phase. Apart from the transfer manipulation, the details of stimulus construction were identical to those in Experiment 1.

Procedure

The procedure was nearly identical to that of Experiment 1. As before, in the training phase of the experiment, participants were simply told that they were going to see a set of items and that they were to study them as carefully as possible, and no information was given as to what would be required from them in

the test phase. In this case, the items were presented on a computer screen, and the display time (10 s) was computer controlled (using the Superlab software for running psychology experiments).

After participants had seen the 20 unique G items presented three times, they were presented with instructions for the test phase. They were told that the training items were all constructed so that they obeyed a complex set of rules and that they had to identify the items that complied with these rules in the test phase, even though the letters, shapes, or cities that would make up the test items were different from the ones that were used in the training phase. In the case of the city stimuli, the routes participants saw in the training phase consisted of American cities, whereas the routes in the test phase consisted of the European ones that were used in Experiment 1. As in Experiment 1, participants were told that the rules were devised to make sure that the company's airplanes were put to best use, and it was added that a European company decided to use exactly the same rules in planning routes between various European cities, so as to justify the mapping from American to European cities. The test items were presented on a computer screen, and participants were asked to indicate whether they thought an item was G or NG by pressing the keys labeled "Good" and "Not-Good" on a computer keyboard.

RESULTS

We proceeded in the same way as in Experiment 1. First, we eliminated participants whose average grammaticality accuracy deviated by 2 standard deviations or more from the overall mean performance. This resulted in one participant being eliminated from the letter condition, two from the shape condition, and one from the city condition. As can be seen in Table 2, grammaticality accuracy in this experiment is near chance with all three kinds of stimuli (performance in transfer AGL experiments is generally lower than in corresponding no-transfer conditions). Grammaticality accuracy was statistically above chance (two-tailed tests, as before) only with the letter and city stimuli, not with the shape stimuli. However, a one-way independent sample ANOVA comparing overall grammaticality accuracy in the three conditions was not significant, $F(2, 53) = .937$, $p = .398$.

We next looked at grammaticality accuracy in each of the possible subsets, as in Experiment 1. As can be seen in Figure 4, the pattern of per-

Table 2. Mean grammaticality accuracy in the three conditions of Experiment 2

Stimuli	Mean	*SD*	*t*-Test against chance
Letters	52.2%	4.2%	$t(18) = 2.3$, $p = .032$
Shapes	51.6%	4.4%	$t(17) = 1.5$, $p = .149$
Cities	53.4%	4.1%	$t(18) = 3.6$, $p = .002$

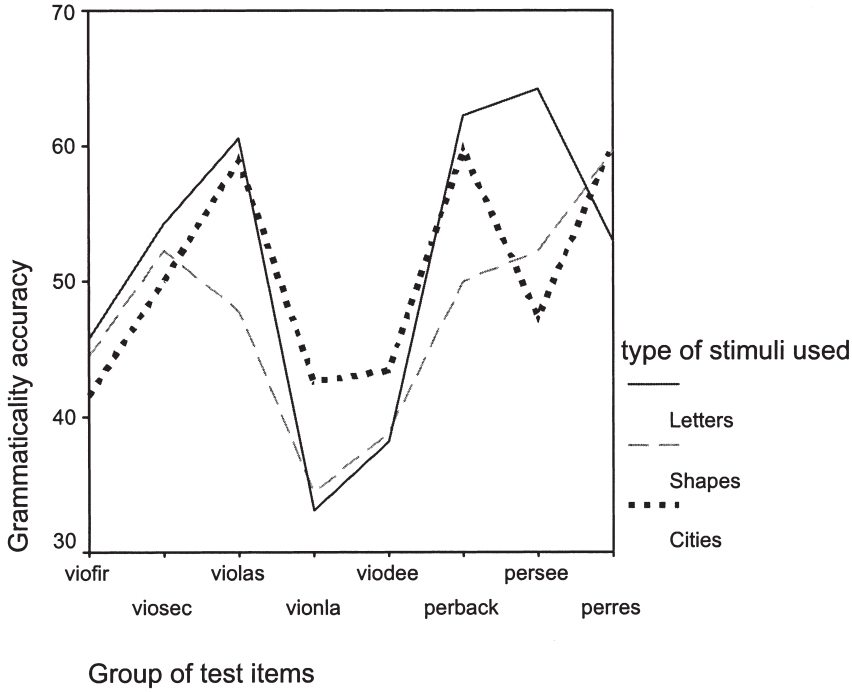

Figure 4. Grammaticality accuracy on different subsets of the test items in Experiment 2. viofir = violations in the first position; viosec = violations in the second position; violas = violations in the last position; vionla = violations in the next to last position; viodee = deep violations; perback = items spelled backwards; persee = G test items repeated from training; perres = novel G test items

formance in the different subsets is nearly identical for the three kinds of stimuli, an intuition that was statistically confirmed with a mixed-design ANOVA, with stimulus type as the between-participants factor and grammaticality accuracy in different subsets of the test items as the within-participants factor (as before, the SPSS Greenhouse–Geisser correction was used), $F(7.9, 211.9) = 1.134$, $p = .342$.

DISCUSSION

Results with the three types of stimuli were remarkably similar and did not support our expectations that the learning process might be impaired with the shape stimuli or the city stimuli. Grammaticality accuracy also was very near chance. However, straightforwardly, inspection of Figure 4 clearly shows that grammaticality accuracy in certain subsets of the test

stimuli was rather high; this observation, together with the single-sample *t* tests against chance, argues against a no-learning null hypothesis.

So far, we have failed to identify conditions in which stimulus appearance might impair the learning process in a way that would reduce overall grammaticality accuracy. As discussed earlier, however, it is likely that the effect of stimulus appearance is subtle and biases the learning process to look for rules regularity or similarity regularity. We tested this intuition in Experiment 3.

EXPERIMENT 3

METHOD

Design

As in Experiments 1 and 2, we created three AGL tasks that varied in how the stimuli appeared (letter strings, sequences of cities, and embedded geometric shapes). We tested 20 University of Ioannina students in each condition. Participants volunteered to take part in the experiment.

Materials

In this experiment we used the Knowlton and Squire (1996) Experiment 1 stimuli, which were based on the finite state language shown in Figure 5. Knowlton

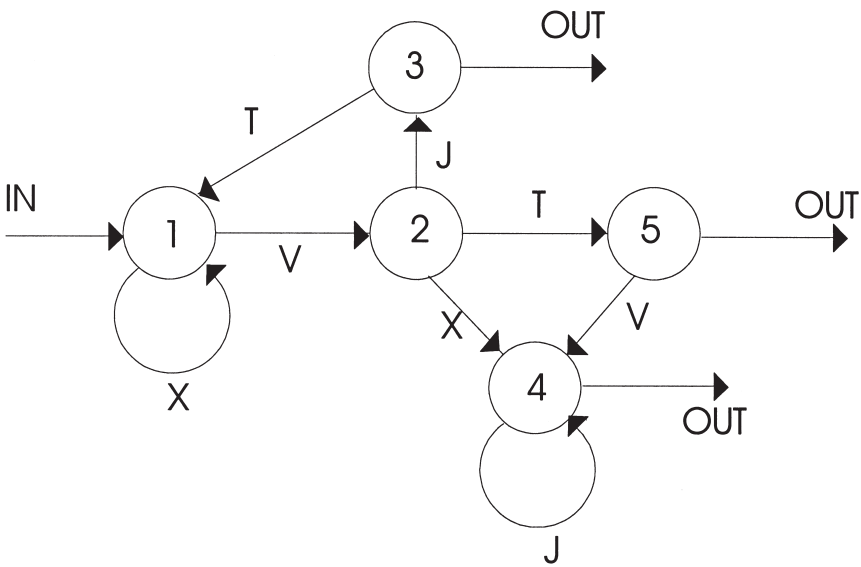

Figure 5. The finite state language used in Experiment 3

and Squire's training set consists of 23 training items and 32 test items, with an equal number of G and NG items. The NG items were constructed from G items on the basis of a single violation (a single change in each NG item would make it G). The test items were selected so that the average global associative chunk strength of the G sequences was the same as the average global associative chunk strength of the NG ones. (In fact, the materials of Knowlton & Squire, 1996, as reported in the Appendix of their article and the later correction to that Appendix, contained a few errors; these were corrected in Pothos & Bailey, 2000, who used the same materials.) This enables a more straightforward interpretation of grammaticality as rule knowledge (see Pothos, 2005b) and also the investigation of grammaticality and global associative chunk strength as two independent factors in participants' performance. As noted earlier, global associative chunk strength can be interpreted as an influence of similarity in performance.

The original Knowlton and Squire (1996) letter strings were used to create city and shape versions of the AGL task in a way analogous to that of Experiment 1. The letters *X, J, V,* and *T* (Figure 5) were mapped to Athens, London, Paris, and Berlin for the city condition. Note that although we used letter strings with the original Latin characters, the city names were translated into Greek. For the shape condition, the letters *X, J, V,* and *T* were mapped to a square, a circle, a diamond, and a hexagon. The dimensions of the stimuli were analogous to those in Experiment 1.

Procedure

We used a procedure similar to that in Experiment 1. Before the training phase, participants were told that they were about to see a set of stimuli composed of letters, cities, or geometric shapes. Participants were asked to observe the stimuli, and they were told that there would be a second phase to the experiment, but they received no other information about what they would have to do with the stimuli or what would be required of them in the second phase. With the city stimuli it was added that the city sequences corresponded to the routes of an airline company. The 23 training items were individually presented three times for 7.5 s each so that no item was presented twice until all items had been presented once (and likewise for the third presentation). After the training phase and before the test phase, participants were told that all the items they had just seen were created on the basis of a complex set of rules, that they were about to see new items, and that they had to decide which of these new items also complied with the rules. It was added that if participants were not sure what these rules were, they should respond on the basis of their intuition. Responses were indicated by pressing a key labeled "yes" (as in "The item complies with the rules") and a key labeled "no" (as in "The item does not comply with the rules") on a computer keyboard. The 32 test items were individually presented once, and they were visible on the computer screen until a response was made. Participants were urged to respond as quickly and accurately as possible, and no corrective feedback was provided. Stimulus presentation and response recording were computer controlled (using the Experimental Run-Time System software for controlling psychology experiments). All instructions appeared in Greek.

RESULTS

As in Experiments 1 and 2, we first eliminated participants whose performance deviated by 2 or more standard deviations from the overall mean performance. In this case, however, we had two indices of performance: grammaticality accuracy, which was computed as the number of G items selected as G in test and the number of NG items rejected as NG; and chunk strength. As in Knowlton and Squire (1996), the test items could be divided into a set of high (global associative) chunk strength and low (global associative) chunk strength; the counterbalancing procedure of Knowlton and Squire means that in the set of high–chunk strength items, the number of G items was the same as the number of NG ones, and likewise for the set of low–chunk strength items. Note that if participants were 100% correct on grammaticality, then their chunk strength performance would be 50% (chance) and, conversely, if they were 100% correct on chunk strength then their grammaticality performance would be 50%.

Given two indices of performance, we eliminated participants whose performance was an outlier either in terms of grammaticality or in terms of chunk strength. In the letter condition, two participants were eliminated, both because of very high performance on grammaticality. In the city condition, two participants were eliminated because of very low performance on chunk strength. Finally, in the shape condition, one participant was eliminated because of very high performance in chunk strength and another because of very low performance on grammaticality. Tables 3 and 4 show mean performance for grammaticality and chunk strength and corresponding single-sample *t* tests (two tailed) against chance. The *t* tests involving chunk strength were all clearly significant, and the tests involving grammaticality were either significant or approaching significance.

With the Knowlton and Squire (1996, Experiment 1) materials it is not possible to look at performance in different subsets of the test items, in the principled way that was possible with Reber and Allen's (1978) test set. However, we can investigate a possible differential effect of stimulus appearance on grammaticality and chunk strength performance. Observing Tables 3 and 4, the following trends are apparent. First, chunk strength performance is higher than grammaticality performance in all three con-

Table 3. Mean grammaticality accuracy in the three conditions of Experiment 3

Stimuli	Mean	*SD*	*t*-Test against chance
Letters	55.4%	9.6%	$t(17) = 2.4$, $p = .029$
Shapes	53.5%	8.2%	$t(17) = 1.8$, $p = .091$
Cities	53.3%	7.1%	$t(17) = 2.0$, $p = .064$

Table 4. Mean chunk strength accuracy in the three conditions of Experiment 3

Stimuli	Mean	*SD*	*t*-Test against chance
Letters	55.7%	7.9%	$t(17) = 3.1$, $p = .007$
Shapes	55.9%	7.9%	$t(17) = 3.2$, $p = .005$
Cities	56.4%	5.4%	$t(17) = 5.0$, $p = .000$

ditions, but this difference is pronounced in the city and shape conditions. By contrast, in the letter condition grammaticality performance is only slightly less than chunk strength performance. We ran a mixed-design ANOVA with condition (letters, shapes, or cities) as a between-participants factor and grammaticality performance versus chunk strength performance as a within-participants factor (the data points in this ANOVA were the average grammaticality performance and average chunk strength performance for each participant). The interaction between the two factors was not significant, $F(2, 51) = .234$, $p = .79$, and neither factor was individually significant. For condition, we found $F(2, 51) = .187$, $p = .83$, and for grammaticality versus chunk strength, $F(1, 51) = 1.30$, $p = .26$. In sum, the differences apparent by inspection of Tables 3 and 4 were not large enough to be statistically supported. Our analyses provide no support for the hypothesis that stimulus appearance might differentially affect grammaticality and chunk strength performance. Note that the observed pattern of results does not change qualitatively if the eliminated outliers are included in the analyses.

DISCUSSION

There is a very intuitive interpretation of the trends observed in Tables 3 and 4. Because the letter stimuli are entirely meaningless and have a highly salient symbolic structure, participants were more inclined to process the stimuli in a way that led to rule knowledge. Similarity biases were evident with the letter stimuli, but the influences of rules and similarity on performance were equivalent. By contrast, the perceptual nature of the shape stimuli presumably made participants attend more to the similarity structure of the training items so that similarity influences were pronounced and rule influences suppressed. Likewise, with the city stimuli participants probably were influenced by their a priori conceptions of what would constitute good airline routes, so they were insensitive to rule regularities and had to resort to a similarity strategy.

None of these suggestions was statistically supported. Although it is clear that stimulus appearance affects the learning process, this influence is

nowhere near enough to identify statistically significant differences. Thus, the tentative conclusion is that stimulus appearance does not alter the balance between rules and similarity influences on the learning process.

GENERAL DISCUSSION

Our aim has been to provide a preliminary investigation of how much the superficial appearance of stimuli affects the learning process. Is it the case that different types of stimuli direct the learning mechanism to operate in different ways? Or is it the case that the learning process will look for stimulus regularity in the same way, regardless of how stimuli look? Our results indicate the latter view to be more accurate.

In evaluating this research, two issues must be addressed. First, there is no limit to the types of stimuli we could examine. We chose to compare AGL performance with stimuli such as letters, cities, and shapes so as to broadly address issues such as the possible influence of general knowledge expectations on learning and whole object processing versus the processing of symbolic sequences. An important limitation in our choices is that with all kinds of stimuli, individual symbols could be identified (even with the shape stimuli, which were meant to give the impression of a whole object, each individual symbol was readily recognizable). With future work we would like to use stimuli in which individual symbols are obscured (cf. Dienes & Longuet-Higgins, 2006; Pothos & Bailey, 2000). Moreover, recent research shows that the learning process may operate in different ways depending on the modality of the stimuli (equivalently, there might be several modality-specific learning mechanisms; Conway & Christiansen, 2005). Accordingly, some of the presently reported manipulations could be very fruitfully applied to comparisons involving learning in different modalities as well.

Second, it is a lot easier to make a case against a null hypothesis rather than in its favor. One can always argue that given enough participants, the trends observed here would become significant. After all, performance was never quite identical in the different conditions we explored. We agree that ultimately it must be the case that the learning process is not totally independent of stimulus format. But how large a difference is large enough to warrant taking into account in AGL methods and theory? In designing the present investigation we adhered very closely to the procedural details of previous corresponding studies. And our results are largely consistent with these studies (e.g., in terms of grammaticality accuracy and finding above-chance performance). Therefore, we claim that our demonstration for AGL performance equivalence regardless of stimulus appearance is as powerful as needed in light of previous AGL research.

In closing, note that establishing the extent to which the learning process in AGL is influenced by stimulus appearance not only has implications for AGL theory but also can increase the utility of AGL in practical contexts. Pothos and Cox (2002) and Pothos and Kirk (2004) applied the AGL paradigm in studies of alcohol abuse and dyslexia, respectively. These applications crucially depend on the extent to which AGL performance with different types of stimuli does not differ. The present findings corroborate the validity of Pothos and Cox's and Pothos and Kirk's methods and hence a perception of the utility of AGL.

Notes

Some of these results were presented at the 1998 annual meeting of the Cognitive Science Society. We would like to thank Pierre Perruchet, John Vokey, and an anonymous reviewer for comments on an earlier version of this manuscript. Additionally, we are grateful to Todd Bailey for his contribution in this project. This research was supported by a European Commission Framework 6 grant, contract 516542 (NEST).

Correspondence about this article should be addressed to Emmanuel M. Pothos, Department of Psychology, Swansea University, Swansea 5A2 8PP, England (e-mail: e.m.pothos@swansea.ac.uk). Received for publication May 20, 2004; revision received April 28, 2005.

References

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 899–912.

Boole, G. (1854). *An investigation of the laws of thought.* New York: Dover.

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: Learning, Memory, and Cognition, 120,* 316–323.

Chan, C. (1992). *Implicit cognitive processes: Theoretical issues and applications in computer systems design.* Unpublished PhD thesis, University of Oxford.

Chater, N., & Hahn, U. (1997). Representational distortion, similarity and the universal law of generalization. *Proceedings of the Similarity and Categorization Workshop, 97,* 31–36.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N., & Miller, G. A. (1958). Finite state languages. *Information and Control, 1,* 91–112.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120,* 235–253.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 24–39.

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

Dienes, Z., & Altmann, G. (2003). Measuring learning using an untrained control group: Comment on R. Reber and Perruchet. *Quarterly Journal of Experimental Psychology, 56A,* 117–123.

Dienes, Z., Altmann, G. T., & Gao, S. (1999). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science, 23,* 53–82.

Dienes, Z., & Longuet-Higgins, C. (2006). Can musical transformations be implicitly learned? *Cognitive Science, 28,* 531–558.

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

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1985). On consciousness in syntactic learning and judgment: A reply to Reber, Allen, and Regan. *Journal of Experimental Psychology: General, 114,* 33–49.

Elman, J. L. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3–71.

Gomez, R. L., & Schvaneveldt, R. W. (1994). What is learned from artificial grammars? Transfer tests of simple association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 396–410.

Hampton, J. A. (1999). The role of similarity in natural categorization. In M. Ramscar, U. Hahn, E. Cambouropoulos, & H. Pain (Eds.), *Similarity and categorization.* Oxford: Oxford University Press.

Higham, P. A. (1997). Dissociations of grammaticality and specific similarity effects in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1029–1045.

Higham, P. A., Vokey, J. R., & Pritchard, J. L. (2000). Beyond dissociation logic: Evidence for controlled and automatic influences in artificial grammar learning. *Journal of Experimental Psychology: General, 129,* 457–470.

Homa, D., & Chambliss, D. (1975). The relative contribution of common and distinctive information on the abstraction from ill-defined categories. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 351–359.

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

Kinder, A., Shanks, D. R., Cock, J., & Tunney, R. J. (2003). Recollection, fluency, and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology: General, 132,* 551–565.

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.

Kruschke, J. K. (1992). ACLOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Malt, B. C. (1994). Water is not H_2O. *Cognitive Psychology, 27,* 41–70.

Marcus, G. F. (2001). *The algebraic mind.* Cambridge, MA: MIT Press.

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

Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1007–1028.

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

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Journal of Experimental Psychology: Perception and Psychophysics, 45,* 279–290.

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., Pacteau, C., & Gallego, J. (1997). Abstraction of covariations in incidental learning and covariation bias. *British Journal of Psychology, 88,* 441–458.

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

Perruchet, P., Vinter, A., Pacteau, C., & Gallego, J. (2002). The formation of structurally relevant units in artificial grammar learning. *Quarterly Journal of Experimental Psychology, 55A,* 485–503.

Pothos, E. M. (2005a). Expectations about stimulus structure in implicit learning. *Memory & Cognition, 33,* 171–181.

Pothos, E. M. (2005b). The rules vs. similarity distinction. *Behavioral and Brain Sciences, 28,* 1–49.

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.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26,* 303–343.

Pothos, E. M., & Cox, W. M. (2002). Cognitive bias for alcohol-related information in inferential processes. *Drug and Alcohol Dependence, 66*(3), 235–241.

Pothos, E. M., & Hahn, U. (2000). So concepts aren't definitions, but do they have necessary *or* sufficient features? *British Journal of Psychology, 91,* 439–450.

Pothos, E. M., & Kirk, J. (2004). Investigating learning deficits associated with dyslexia. *Dyslexia, 10,* 61–76.

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

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

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functional interpretation. *Cognition, 6,* 189–221.

Reber, R., & Perruchet, P. (2003). The use of control groups in artificial grammar learning. *Quarterly Journal of Experimental Psychology, 56A,* 97–115.

Redington, F. M., & Chater, N. (1996). Transfer in artificial grammar learning: Methodological issues and theoretical implications. *Journal of Experimental Psychology: General, 125,* 123–138.

Rips, L. J. (1989). Similarity, typicality and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning.* Cambridge: Cambridge University Press.

Rips, L. J. (2001). Necessity and natural categories. *Psychological Bulletin, 127,* 827–852.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Saffran, J. R., Aslin, R. N., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

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., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology, 50A,* 216–252.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammar. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 328–344.

Whittlesea, B. W., & 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., & 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.