Syntactic transfer in artificial grammar learning

T. BEESLEY
University College London, London, England

A. J. WILLS
University of Exeter, Exeter, England

AND

M. E. LE PELLEY
Cardiff University, Cardiff, Wales

In an artificial grammar learning (AGL) experiment, participants were trained with instances of one grammatical structure before completing a test phase in which they were required to discriminate grammatical from randomly created strings. Importantly, the underlying structure used to generate test strings was different from that used to generate the training strings. Despite the fact that grammatical training strings were more similar to nongrammatical test strings than they were to grammatical test strings, this manipulation resulted in a positive transfer effect, as compared with controls trained with nongrammatical strings. It is suggested that training with grammatical strings leads to an appreciation of set variance that aids the detection of grammatical test strings in AGL tasks. The analysis presented demonstrates that it is useful to conceptualize test performance in AGL as a form of unsupervised category learning.

In a typical artificial grammar learning (AGL) task, participants are first asked to memorize a set of letter strings (e.g., XMTRV, XTVMRX). Following this incidental-training phase, participants are told that the letter strings that they have memorized were created from a grammatical structure but are usually not informed about the specific rules of the structure. In order to test whether learning occurred during the incidental-training phase, the participants are asked to discriminate novel grammatical strings from nongrammatical strings. Participants can perform this discrimination (see A. S. Reber, 1993, for a review), and the poor verbal expression of knowledge exhibited by participants has led to the suggestion that learning in the AGL task occurs implicitly (e.g., A. S. Reber, 1993; but see Shanks & St. John, 1994).

Many researchers have examined the representational properties of the knowledge acquired in the AGL task (for a review, see Pothos, 2007). A large body of evidence has suggested that during the training phase, participants *chunk* strings into fragments to aid memorization (e.g., Perruchet & Pacteau, 1990). Grammaticality judgments at test are then assumed to be based on the fragment content of the strings; test strings containing a higher proportion of fragments encoded during training will have a higher probability of being endorsed as grammatical (Perruchet, Vinter, Pacteau, & Gallego, 2002). Furthermore, participants might learn a host of *microrules* relating to what makes strings grammatical (Dulany, Carlson, & Dewey, 1984). These might, for instance, consist of the simple

distributional statistics of grammatical strings: information about string length or legal starting and ending letters (see Shanks, Johnstone, & Staggs, 1997).

Many studies have suggested that item similarity contributes to test string discrimination. For instance, Vokey and Brooks (1992) demonstrated that the similarity between a test string and a particular training string will influence grammaticality decisions independently of a main effect of grammaticality (i.e., its similarity to the complete set of training items). Pothos and Bailey (2000) showed that the generalized context model (Nosofsky, 1986), an exemplar model of categorization, can also explain variations in test performance during an AGL task. Such simulations imply a parallel between AGL and categorization processes; grammaticality decisions in AGL are assumed to be based on the average similarity between a test string and all members of the grammatical training set, so that category membership decisions (grammatical or nongrammatical) map directly onto grammaticality decisions. Given the overwhelming support for category formation under incidental-learning conditions (see Smith, 2008), models of categorization undoubtedly provide a valuable context in which to discuss AGL.

The present article examines an additional source of information that may be used in AGL—namely, information about set variance. Since all grammatical strings are created from a particular underlying structure, the variance among grammatical strings in any particular subset is likely to be low. In contrast, randomly created non-

T. Beesley, t.beesley@ucl.ac.uk

grammatical strings do not share an underlying structural basis, and so, variation among the strings in this set will be higher. How would knowledge of set variance facilitate performance in an AGL task? Performance in AGL tasks is, in part, driven by learning during the discrimination test phase (e.g., R. Reber & Perruchet, 2003; Redington & Chater, 1996). Thus, the discrimination test involves not only the process of generalization from memorized strings, but also the development of a new representation of the grammatical structure during the test phase. One possibility, therefore, is that training with grammatical strings generates knowledge regarding the expected level of variance among a set of grammatical strings, which can then be used to produce above-chance classification of grammatical strings during the test phase.

In order to test this possibility, we adapted a condition from A. S. Reber (1969), in which participants were trained with strings from one grammatical structure before being transferred to a second phase containing strings from a *novel* grammatical structure. A. S. Reber (1969) found that participants showed positive transfer in this condition; performance was better for the second grammatical structure than for the first.[1] One could view this as a demonstration of *syntactic transfer*—a positive transfer effect from learning about an initial grammatical structure to learning about a second, novel, grammatical structure. A demonstration of syntactic transfer might suggest that participants have acquired some nonspecific knowledge about grammatical sets, such as an appreciation of the variation among grammatical strings. However, it is unclear whether the facilitated performance in A. S. Reber's (1969) second phase truly reflected transfer of knowledge about *grammaticality* or whether it simply reflected increased practice with procedural aspects of the task (e.g., memorization, recall). Furthermore, A. S. Reber's (1969) grammatical structures shared many common features, and therefore, transfer performance may have reflected generalization of item-specific knowledge from training strings, rather than the more abstract form of transfer considered here. In the present experiment, we reexamined syntactic transfer, using grammatical structures designed to control for item-specific factors known to affect AGL performance.

In designing the grammatical structures used here, our aim was to limit the potential for item-specific knowledge acquired during the training phase to affect grammaticality decisions at test. Therefore, the two grammatical structures (used for training and test) did not share any bigram or trigram information (chunks of two and three letters). If participants' process of memorization during the training phase involves memorization of chunks, the absence of such fragments from the grammatical test strings means that this information could not be used to classify these strings as grammatical. Furthermore, nongrammatical strings contained an equal proportion of chunks from the two grammatical structures. Hence, purely in terms of fragment information, grammatical training strings were more similar to nongrammatical test strings than they were to grammatical test strings (since grammatical and nongrammatical strings shared a small, but nonzero, proportion of fragments, whereas strings from the two different

grammars shared zero fragments; see the Method section). Consequently, if participants were to rely on knowledge of string fragments to make grammaticality decisions during the test phase, nongrammatical strings, rather than grammatical strings, would be more likely to be endorsed as grammatical. In contrast, if positive transfer were observed in the present experiment (i.e., an ability to discriminate grammatical and nongrammatical test strings in participants trained on grammatical strings, but not in participants trained on nongrammatical strings), this would provide evidence for an abstract form of syntactic transfer.

METHOD

Participants and Apparatus

Fifty-six undergraduates participated for payment of £4; 28 were assigned to the *grammatical trained* (GT) condition, and 28 to the *nongrammatical trained* (NGT) condition. Testing was conducted using PCs.

Design

Figure 1 shows the two grammatical structures (Grammar A and Grammar B; hereafter, GA and GB). Both structures used the letters M, R, X, T, and V. Grammatical strings were created by passing through the structure, following the direction of the arrows between nodes. An example string from GA is MRVRV. The two grammatical structures are distinct in several ways. First, the grammars have unique starting and ending letters. Second, they create sets of strings that are unique in their *bigram* content (pairs of letters; e.g., MM, XT). Each bigram could feature in strings created by only one of the two grammatical structures. By extension, the structures therefore also contained no overlapping chunks of any greater length. Of the 25 possible bigrams, 12 were assigned to GA and 13 to GB. For each set of 5 bigrams starting with the same letter, 2 were used in one grammatical structure and 3 in the other. For repetition bigrams (e.g., MM), 2 were assigned to GA and 3 to GB.

A large pool of strings was created for each grammar, from which strings were selected at random until each of the two selected sets of grammatical strings had the following properties: (1) Exactly half of the grammatical strings in each set contained a repetition bigram, and (2) no more than one repetition occurred in each string. String length was limited to between four and seven letters.

Nongrammatical strings were created by randomly selecting letters for strings with lengths of four to seven. No more than one repetition bigram could appear in a nongrammatical string. From a large pool of nongrammatical strings, those strings containing bigrams from both GA and GB were selected for use in the experiment. Nongrammatical strings with lengths five and seven contained an equal proportion of GA and GB bigrams. For strings with lengths four and six, this was not possible (such strings comprise three and five bigrams, respectively), and so, for these strings, half contained one more GA bigram, and half one more GB bigram. All of the string-generation procedures above were used so as to make it difficult to account for a positive transfer effect by similarity-based processes (e.g., exemplar/fragment encoding).

For encoding of variability to be a plausible means by which syntactic transfer could occur, set similarity (the inverse of set variability) must be higher for grammatical sets of strings than for nongrammatical sets of strings. In order to verify this property, we estimated these similarities using the Pearce (1987) similarity equation:

$$_aS_b = \frac{\text{com}}{N_a} \cdot \frac{\text{com}}{N_b},$$

where $_aS_b$ is the similarity of items a and b, N_a and N_b are the number of features in stimuli a and b, and com is the number of features common to a and b. We instantiated two different versions of this model. In the Pearce (letter) version, individual letters of the strings
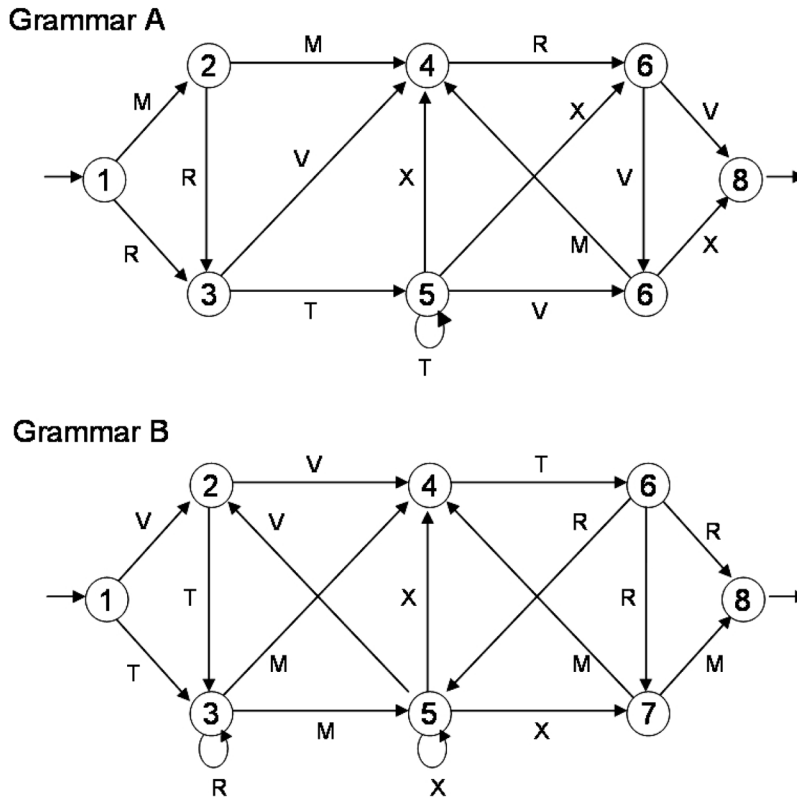
Figure 1. The two grammatical structures used. Grammatical strings are created by making a pass from Node 1 through to Node 8, in the direction of the arrows between nodes.

constituted position-specific features. For example, on this basis, the strings TMXMTR and TRRMXM have two common features (T in Position 1 and M in Position 4) and have a similarity of .111. In the Pearce (bigram) version, bigrams constituted the position-specific features. For example, strings MRTXRVX and MRVRVX have one common bigram (MR in Bigram Position 1) and have a similarity of .033. Both metrics range between 0 (no overlap) and 1 (identity). We also calculated associative chunk strengths (ACS) for each set of test strings (Knowlton & Squire, 1994). This is done by first calculating the frequency with which a bigram or trigram occurs in the set of training strings. These chunk frequencies are then summed for all chunks within a test string and averaged to provide the associative chunk strength of the test string.

Table 1 shows the mean within-set and between-set similarities, using the Pearce equation, and the mean ACS for sets of test strings as a function of the training strings presented. These analyses illustrate that the two sets of grammatical strings are less similar to each other than each of them is to the set of nongrammatical test strings. Furthermore, within-set similarities (and ACS) were higher for the sets of grammatical strings than for the sets of nongrammatical strings. There was a high correlation between ACS values for test strings and appropriate Pearce similarity values (i.e., the mean similarity of each test string to the entire set of training strings), using either the letter [Pearson's $r(320) = .82$] or the bigram version [$r(320) = .86$] of the Pearce metric.

Table 2 shows descriptive statistics for the sets of strings; the statistics of all four sets are matched as closely as possible. All the strings are presented in the Appendix.

Procedure

GT participants were trained on one grammatical set of strings before a test phase containing strings from the *untrained* grammar

(which was not presented during the training phase) and nongrammatical strings (NG_2). The assignment of GA and GB to training and test grammars was counterbalanced in condition GT. NGT participants received training with NG_1 strings, before receiving a test phase with grammatical strings and NG_2 strings. The use of GA and GB in the test phase was counterbalanced for the NGT condition. The assignment of response keys ("z" and "m" keys) to classification responses in the test phase was counterbalanced within each subcondition.

The participants were initially shown example letter strings and were instructed that their task was to memorize a set of strings as best they could. In the training phase, strings were presented sequentially (one every 5 sec), until all 20 were displayed simultaneously in one column on-screen. The participants then received instructions regarding the nature of structured strings and details of the test phase. GT participants were informed that they had been memorizing structured strings. NGT participants were informed that they had been memorizing randomly created strings. All the participants were then told that they would be asked to try and discriminate between grammatical and nongrammatical strings in a test phase, that structured strings were created from a set of rules governing the order of letters, and that these rules stated which letters could follow and precede other letters. GT participants were informed that the grammatical strings in the test phase were created from a grammatical structure different from that which created their training strings. All the participants were told that this task was difficult, that they were not expected to make 100% correct classifications, that they should not be concerned if they acted on impulse, and that half of the test strings were grammatical and half were nongrammatical.

Before the test phase, the participants completed a key-training phase in which they practiced responding to the words "grammatical" and "nongrammatical" with the appropriate keys, with feed-

Table 1
Within- and Between-Set Similarity Results Using the Pearce Equation With Individual Letters
As the Features and With Bigrams As Features and Mean Associative Chunk Strength (ACS)
of Test Strings As a Function of Training Strings

Training Strings	Pearce (Letters)				Pearce (Bigrams)				ACS of Test Strings			
	GA	GB	NG_1	NG_2	GA	GB	NG_1	NG_2	GA	GB	NG_1	NG_2
GA	.226	.032	.055	.059	.164	0	.006	.008	*8.49*	0	*2.59*	2.65
GB		.211	.058	.056		.160	.010	.007	0	*8.81*	*2.25*	2.34
NG_1			.150	.053			.107	.008	2.33	2.32	*3.51*	2.52
NG_2				.142				.102	*2.53*	*2.30*	*2.52*	*3.48*

Note—GA and GB are grammatical sets of strings produced by the grammars shown in Figure 1. NG_1 is the nongrammatical set of training strings. NG_2 is the nongrammatical set of test strings. See the text for details of computations. Mean ACS data are calculated on the basis of chunk frequency obtained from the training strings. Empty cells for the two versions of the Pearce equation are a result of symmetry (e.g., GA_GB = GB_GA). ACS values presented in italics relate to conditions that did not occur in the experiment (e.g., NG_1 was never presented at test; no participants who had received GA strings during training experienced GA strings on test; etc.). These values are included for completeness.

back provided after each response (e.g., "correct response"). Ten "grammatical" and 10 "nongrammatical" trials were presented in random order.

The test phase contained 80 trials, two presentations of each of the 40 strings. Each string was repeated only after all the test stimuli had been presented once. For each participant, each half of the test phase involved a randomized ordering of the test strings. Strings from the last 3 trials of the first presentation half could not be presented on the first 3 trials of the second half. Each subblock of 10 strings contained exactly 5 grammatical strings.

The participants were asked to respond to each string, using the appropriate key according to whether they thought the string was grammatical or nongrammatical, after which "G" or "NG" appeared next to the string according to the response made (with no feedback on accuracy provided). The next stimulus appeared after a response–stimulus interval of 1 sec. If the participants failed to respond within 3 sec, "T/O" (time-out) was presented next to the string and was recorded as such. Once made, responses could not be changed. Items appeared successively in two columns, with 20 strings appearing in each column. Once the first 40 strings had been presented, the screen was cleared for the second presentation.

RESULTS

Three GT participants and 3 NGT participants produced accuracy below 75% on the simple key-training phase. These 6 participants were removed from subsequent analyses, since they were unable to use the response mappings appropriately. The overall pattern of data was the same with these participants included. Significance in all the analyses reported below was assessed against $\alpha = .05$.

The mean percentage of trials on which a timeout was recorded was low in both the GT (1.7%) and NGT (1.2%) conditions. GT participants made significantly more correct responses than did NGT participants [54.3% vs. 49.8%; $t(48) = 2.21$]. Figure 2 shows grammatical endorsement rates for grammatical and nongrammatical strings during the test phase (i.e., hits vs. false alarms), for both conditions. A signal detection analysis[2] confirmed that GT participants were significantly better at discriminating between test items than were NGT participants [$d' = 0.22$ and -0.01, respectively; $t(48) = 2.22$]. Discrimination was significantly better than chance in the GT condition [$t(24) = 3.09$], but not in the NGT condition ($t < 1$). There was no difference in response bias between the GT and NGT conditions [$\beta = .99$ and 1.00, respectively; $t(48) = 1.02$, $p = .31$].

These results suggest that the NGT participants were unable to accurately classify strings during the test phase. One possibility is that the NGT participants were simply responding randomly at test. Alternatively, the NGT participants might have been able to categorize strings into grammatical and nongrammatical sets but, without any experience of grammaticality during the training phase, might have been unable to assign appropriate category labels (*grammatical* and *nongrammatical*) to these categories. These two possibilities can be distinguished by using an absolute value of d' (hereafter, d'_{abs}; see, e.g., Haslam et al., 2007, for a related approach in the study of unsupervised classification), where

$$d'_{abs} = |z(\text{hits}) - z(\text{false alarms})|.$$

Thus d'_{abs} is the absolute value of the difference in the z transform of the hit rate and false alarm rate.

If the inability to assign category labels consistently were indeed behind the poor performance in condition NGT, d'_{abs} scores should be similar in the NGT and GT conditions. However, if the poor performance in condition NGT reflected poor discrimination of the two sets of test strings, d'_{abs} would reveal lower levels of consistency in group NGT. Analysis of d'_{abs} produced results consistent with the former hypothesis: GT and NGT conditions did not differ significantly ($d'_{abs} = 0.33$ and 0.30, respectively; $t < 1$).

Since the minimum value of d'_{abs} is zero, the value expected by chance will be greater than zero. Monte Carlo

Table 2
Descriptive Statistics for the Sets of Letter Strings

Set	Length		Proportion of Repetitions		Proportion of Unique Letters	
	M	SD	M	SD	M	SD
GA	5.85	1.04	.50	.51	.72	.11
GB	5.85	0.88	.50	.51	.66	.11
NG_1	5.84	1.02	.45	.51	.65	.10
NG_2	5.80	1.11	.50	.51	.67	.14

Note—GA and GB are grammatical sets of strings produced by the grammars shown in Figure 1. NG_1 and NG_2 are nongrammatical sets of strings. Statistics shown are mean length, mean proportion of strings containing repetitions (e.g., TT), and mean proportion of unique letters per string, for the four sets of strings.
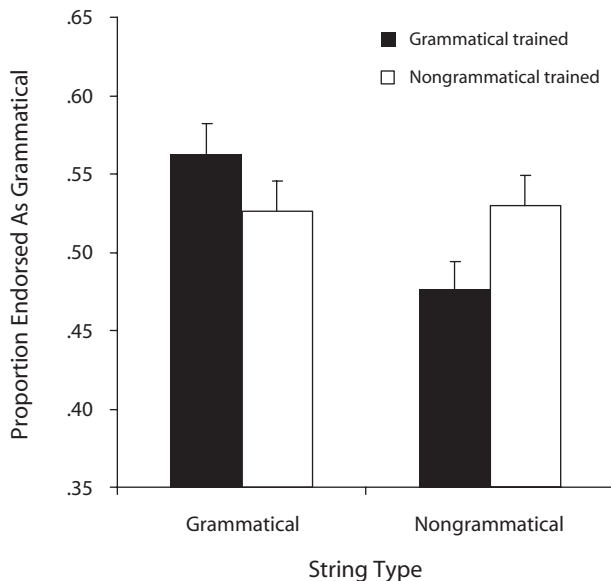
Figure 2. Proportions of grammatical (test grammar) and nongrammatical strings endorsed as grammatical for participants trained with grammatical strings (training grammar) and participants trained with nongrammatical strings.

methods (100,000 iterations) indicated that for an unbiased random responder (i.e., one for whom the probability of responding *grammatical* on any given trial is always .5), the mean value for d'_{abs} is 0.227, and the 95th percentile of the distribution is 0.286. Hence, both GT and NGT participants showed greater consistency, as measured by d'_{abs}, than would be expected by chance. The GT and NGT conditions did not differ on an absolute measure of bias, $\beta_{abs} = \max(\beta, 1/\beta)$ (1.04 and 1.03, respectively; $t < 1$).[3]

Although our sets of grammatical and nongrammatical strings were generated so as to rule out many alternative interpretations of the main transfer effect, certain elements were not perfectly balanced between all sets of strings. For instance, in both sets of grammatical strings, the first and last letters of each string were always different; this was not always the case for nongrammatical strings. The frequency of an alternation (e.g., MXM) was also greater for nongrammatical (55%) than for grammatical (30%) strings. Multiple regression analysis (cf. Lorch & Myers, 1990) with predictor variables of grammaticality and the two additional factors noted above (first/last letter match and alternation) revealed grammaticality as the only variable that significantly predicted endorsement rate [$t(56) = 2.44$]. Stepwise regression produced a model that included grammaticality but excluded both of the other factors [$F(1,58) = 6.52$].

DISCUSSION

The present data provide a clear demonstration of syntactic transfer in AGL. The participants trained on grammatical strings classified strings from a novel grammatical structure as grammatical at above-chance levels. Given the design of the grammatical structures used to generate

training and test strings, it seems unlikely that this grammaticality effect could have resulted from the direct transfer of item-specific knowledge from training to test. Our account of these data is that accurate test classification relied on a more abstract aspect of participants' knowledge: an acquired appreciation of the level of variance within a set of strings described as *grammatical*. When faced with a discrimination phase involving strings from a new grammatical structure and nongrammatical strings, the participants in both conditions demonstrated an equivalent ability to distinguish between the high-variability (nongrammatical) and low-variability (grammatical) sets of strings, as shown by the above-chance levels of *consistency* (i.e., the ability to consistently classify test strings into two distinct sets). The above-chance consistency scores demonstrate an ability to pick up on the differences between strings during the test phase alone (cf. Redington & Chater, 1996). This ability may reflect sensitivity to the repeating fragments in the grammatical test strings, akin to the *internal chunk strength* sensitivity suggested by Meulemans and Van der Linden (1997). However, our data demonstrate a novel difference between participants trained on grammatical and nongrammatical strings: Only the participants who had acquired knowledge about grammaticality during the training phase could identify strings with low variability (high internal chunk strength) as the grammatical set in the test phase, and therefore, only this condition produced above-chance levels of *accuracy*. Thus, these data suggest that the detection of structured material is enhanced by an appreciation of grammatical set variance acquired during AGL (see Flannagan, Fried, & Holyoak, 1986, for a related phenomenon in perceptual classification learning). The specific processes by which this appreciation of set variance is acquired are not addressed by the present data.

An alternative interpretation of the present data is that training with nongrammatical strings led to an impairment in a preexisting ability to accurately assign category labels at test, rather than the facilitation of this ability in the grammatically trained group. The present data do not allow us to decide between these two alternatives. Future work may benefit from the use of an untrained control group, which could be compared with grammatical-trained and nongrammatical-trained groups to assess facilitation and impairments caused by training, respectively (cf. R. Reber & Perruchet, 2003).

The present data have implications for other transfer effects that have been studied extensively in the AGL paradigm. For instance, in *symbolic transfer* procedures (e.g., Altmann, Dienes, & Goode, 1995), grammatical test items are created from the same structure as that used to generate training items but are presented in a different modality (e.g., training may involve auditory stimuli, whereas the test phase may involve visual stimuli). Despite this difference in the surface features of training and test stimuli, participants can still discriminate grammatical from nongrammatical test items. Although the mechanisms underlying symbolic transfer are not fully understood, current theories rely on a process of abstraction between stimulus encoding and test performance (see Lotz & Kinder, 2006).

We suggest that the present data also speak to theories of symbolic transfer: Participants trained with grammatical strings will approach the test phase with a greater appreciation of the level of variance among grammatical strings than will control participants. Future work will be needed to explore the relationship between symbolic and syntactic transfer effects.

REFERENCES

ALTMANN, G. T. M., DIENES, Z., & GOODE, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 899-912.

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.

FLANNAGAN, M. J., FRIED, L. S., & HOLYOAK, K. J. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 241-256.

HASLAM, C., WILLS, A. J., HASLAM, S. A., KAY, J., BARON, R., & MCNAB, F. (2007). Does maintenance of color categories rely on language? Evidence to the contrary from a case of semantic dementia. *Brain & Language*, **103**, 251-263.

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

LORCH, R. F., JR., & MYERS, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 149-157.

LOTZ, A., & KINDER, A. (2006). Transfer in artificial grammar learning: The role of repetition information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 707-715.

MEULEMANS, T., & VAN DER LINDEN, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1007-1028.

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

PEARCE, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, **94**, 61-73.

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., 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. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, **133**, 227-244.

POTHOS, E. M., & BAILEY, T. M. (2000). The role of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 847-862.

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

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

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

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

SHANKS, D. R., JOHNSTONE, T., & STAGGS, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology*, **50A**, 216-252.

SHANKS, D. R., & ST. JOHN, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral & Brain Sciences*, **17**, 367-447.

SMITH, E. E. (2008). The case for implicit category learning. *Cognitive, Affective, & Behavioral Neuroscience*, **8**, 3-16.

VOKEY, J. R., & BROOKS, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 328-344.

NOTES

1. In A. S. Reber's (1969) task, participants recalled memorized letter strings to criterion. Learning was measured by the number of trials to criterion.

2. For all comparisons presented in this article, the same pattern of results was found when an analysis was conducted using percent correct.

3. Due to a coding error, the Grammar B string VTMVVTR was presented to participants as VTVVTR; this contains the bigram TV, which is permissible in Grammar A. An alternate analysis, excluding all test strings containing the bigram TV, leads to the same conclusions for d', d'_{abs}, β, and β_{abs} as the analysis presented in the main text.

APPENDIX
Letter Strings Used in the Experiment

GA	GB	NG$_1$	NG$_2$
MMRV	TMXMTR	MRRXVTX	MTXRXV
MMRVX	TMXTRXM	MTMTVM	MVRV
MRTTVX	TMXXM	MXMRT	MVXRTMV
MRTTXRV	TMXXMTR	MXRT	MXVXT
MRTTXVX	TRMTR	RTMVR	RMTTVT
MRTVMRV	TRMTRM	RVMVT	RRVXM
MRTVX	TRMXM	RXRV	RRXRV
MRTXRVX	TRMXXTR	RXVXRR	RVTVXTR
MRVRV	TRRMTR	TRTRTX	TMMTV
MRVRVX	TRRMXM	TTMVTXR	TMRX
RTTVMRV	VTMTR	TVMVVTX	TMVRVMX
RTTVX	VTMTRM	TVVMX	TTRTXT
RTTXRVX	VTMXM	VMRXVV	VRMVR
RTTXV	VTMXTR	VMTXT	VRVTRT
RTTXVX	VTMXTRM	VRMTVMT	VTTM
RTVMRV	VTMXXM	VRRMTX	VXXTVT
RTVMRVX	VTRRMTR	XMRMXVR	XRRTMXV
RTXVX	VTMVVTR	XMTXTVR	XTVRXXR
RTXRVX	VVTR	XRXVVRM	XVRTMTM
RTXVMRV	VVTRM	XVTVTR	XXTRVXR