Unconfounding Similarity and Rules in Artificial Grammar Learning

Todd M. Bailey (todd.bailey@psy.ox.ac.uk)
University of Oxford; Department of Experimental Psychology
South Parks Road, Oxford OX1 3UD
United Kingdom

Emmanuel M. Pothos (pothos@psy.ox.ac.uk)
University of Oxford; Department of Experimental Psychology
South Parks Road, Oxford OX1 3UD
United Kingdom

Abstract

Artificial grammar learning provides a principled experimental framework to investigate the roles of similarity and rule-induction mechanisms in category generalisation. Past attempts to disentangle these two mechanisms may be criticised for employing insensitive measures of similarity with little theoretical or empirical motivation, for failing to achieve independent measures of the effects of similarity and rule-induction components, and, with several notable exceptions, for confining stimuli to the domain of letter strings. The present work reports on two studies of artificial grammar learning using a standard grammar to arrange nested geometric shapes (Experiment 1) and angles between connected lines (Experiment 2). Grammaticality judgements for novel items are significantly above chance in both experiments. Similarity judgements for pairs of stimuli are used as the basis for modelling grammaticality judgements, using an exemplar-based model of categorisation. We test for independent contributions of similarity and rule-induction mechanisms by fitting nested regression models. Similarity is significant in accounting for grammaticality judgements in both experiments. Rule-induction has an additional, independent effect in Experiment 2, but not in Experiment 1. We discuss the implications of these results and their relationship to previous studies.

Introduction

Having seen a few instances from a given category, we readily generalise our experience to classify new instances as likely or unlikely to be members of the same category. This sort of inductive inference has been a puzzle for philosophers and psychologists alike, in part because there is no logically ideal inference algorithm to use as a benchmark for human performance (Goodman, 1954; Watanabe, 1985). Two broad classes of knowledge have been hypothesised to explain generalisation behaviour: memory of specific instances or exemplars, and summary information abstracted or compiled across relevant instances (cf. Brooks & Vokey, 1991; Hahn, 1996; Hahn & Chater, 1998). Knowledge of exemplars could be generalised by comparing a new instance with memory traces of previous instances, and classifying the new instance according to its similarity with old instances. Summary knowledge, on the other hand, might take the form of classification rules induced from previous instances (e.g. a generative grammar, or necessary and sufficient conditions on category membership), a single prototype representing the most typical case, or statistical profiles of instance parts. A fundamental question in cognitive psychology is whether productive use of knowledge involves summary data, or reference to memories of specific instances. For convenience, we will refer to summary knowledge as rules, though we emphasise our intention merely to contrast an exemplar-based account of generalisation with other accounts based on summary data of any sort.

Reber (1967) showed that after studying a set of strings generated by an artificial grammar, participants could discriminate between new strings that complied with the rules of the grammar and strings that violated those rules. Reber suggested that participants were learning about the abstract rule structure of the grammar that was used to generate the grammatical strings. But is an abstract representation of the grammar required to achieve above chance performance on such artificial grammar learning (AGL) tasks? Dulany, Carlson, and Dewey (1984, 1985) argued that participants acquire "correlated grammars," that is a set of "microrules" which approximate the true grammar, but might at the same time include unrepresentative or even wrong rules. Dulany et al.'s theory can be seen as a rule-based account of AGL, though the knowledge acquired according to this account is fragmentary in nature. Perruchet and Pacteau also suggested that participants acquire fragmentary knowledge, though instead of microrules Perruchet and Pacteau suggest that participants learn which bigram fragments occur in the training set (Perruchet & Pacteau, 1990; Perruchet, 1994; Perruchet, Gallego, & Pacteau, 1992). They showed that participants were more likely to make errors with strings that contained legal bigrams in illegal positions compared to strings that contained illegal bigrams (though cf. Gomez & Schvaneveldt, 1994; Redington, 1996).

Vokey and Brooks (1992, 1994; Brooks, 1978; Brooks & Vokey, 1991) proposed that grammaticality decisions are driven not by adherence to the rule-structure of the finite state language employed, but by the similarity of a test item to the training items. Vokey and Brooks measured similarity by counting the number of letters different between two strings. They reported significant effects of both grammaticality and similarity, that is, items were more likely to be endorsed as grammatical if they were actually grammatical, but also if they were more similar to the training items. Instead of counting shared letters, Knowlton and Squire (1996, Exp.1) assessed similarity by calculating "chunk strengths" (Servan-Schreiber & Anderson, 1990; also Servan-Schreiber, 1991), which measure bigram/trigram overlap between test and training items. Knowlton and Squire argued that if similarity and grammaticality are both significant in predicting participants' grammaticality judgements, both rule-based and exemplar-based learning must be taking place. One weakness in their argument is that their measure of similarity is simply assumed a priori, not grounded empirically in the perceptual similarity of their stimuli. Unless the model of similarity rests on firm footings it is a poor benchmark against which to assess the relative contributions of similarity and rules. Some alternative measure of similarity can always be devised which accounts for grammaticality judgements without appealing to additional knowledge of grammatical rules (cf. Redington, 1996). Redington (1997), for example, showed that a single fragment-based procedure can account for both the similarity and the grammaticality results of Knowlton and Squire (1996). Finally, the statistical analysis employed by Knowlton and Squire does not unambiguously support the conclusion that both rule-based and exemplar-based learning took place, because it fails to unconfound the partially correlated factors of similarity and grammaticality.

In this work we assess the effects of similarity empirically in terms of ratings on the items involved in an AGL task. To facilitate comparison with other studies, our stimuli are derived by applying simple mappings to the stimuli used by Knowlton and Squire (1996). The mappings produce abstract visual figures, while preserving the grammatical structure of the original stimuli. The move away from letter strings emphasises the fact that the type of learning seen in AGL tasks is by no means confined to tasks involving letter strings (cf. Pothos & Chater, 1997). Also, we hoped that using abstract visual figures would provide a more natural set of stimuli for eliciting similarity judgements. Similarity ratings were used to derive a spatial representation of these items through a multidimensional scaling (MDS) procedure, based on Shepard's theory of psychological spaces (Shepard, 1980, 1987; Nosofsky, 1992). In this way, instances are encoded in a multidimensional space which preserves the relative similarities between items. Classification performance is modelled using the generalised context model of categorisation (GCM) to fit instances in the psychological space to grammaticality endorsements (Medin & Schaffer, 1978; Medin, 1986; Nosofsky, 1991, 1990, 1989, 1988a, 1988b). The GCM assumes that new instances are classified depending on how similar they are to previous instances of various categories.

Our choice of the GCM was motivated by general theoretical considerations. The model has provided excellent fits in a variety of studies comparing it with other models (Nosofsky, 1991, 1990, 1989, 1988a, 1988b). Also, since the GCM is equivalent to a non-parametric optimal classification boundary estimator, if similarity to training items had any influence on classification, GCM would be able to identify it (Ashby & Alfonso-Reese, 1995; McKinley & Nosofsky, 1995).

Experiment 1

Participants

All 16 participants except one were University of Oxford students who received five pounds for taking part in the study. The experiment lasted for approximately an hour and fifteen minutes.

Materials

The artificial grammar was identical to the one used by Knowlton and Squire (1996; Exp.1), and is shown in Figure 1. The 23 training strings and 32 test items were constructed by mapping the letters in Knowlton and Squire's material, V, X, J, and T, to a circle, a hexagon, a square, and a diamond, respectively (for an example, see Figure 2). Geometric shapes corresponding to later letters within a string enclosed shapes corresponding to earlier letters.

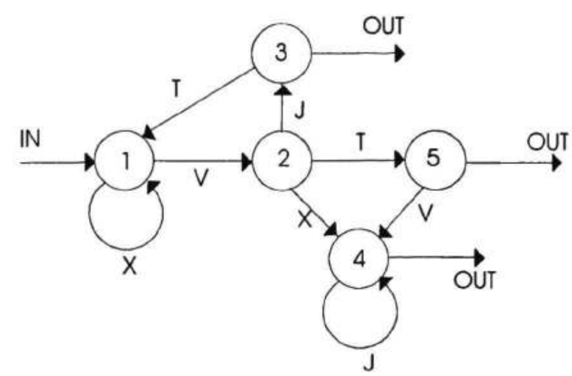

Figure 1: The finite state language (Knowlton & Squire, 1996; Exp.1).

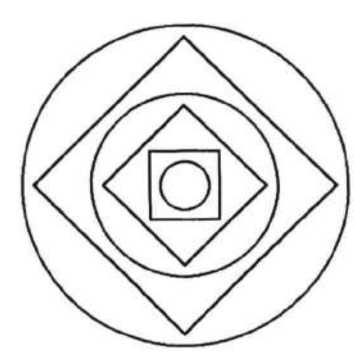

Figure 2: Sample stimulus used in Experiment 1, corresponding to the string VJTVTV.

Procedure

In the first part of the experiment participants viewed the training items on a computer screen for five seconds each. The entire set of training items was presented three times in different random orders.

After the end of the training session, participants were told that the order of the geometric figures in the training items had been determined by a complex set of rules. Participants were asked to classify new items according to whether each item was consistent (grammatical) or inconsistent (ungrammatical) with the rules. The test presented the 32 test items twice in different random orders. No feedback was provided.

In the third part of the study, participants rated the similarity between pairs of stimulus items, on a scale from 1 to 9 (Shin & Nosofsky, 1992). Each trial involved a central fixation point for 250 ms, followed by one item, then another fixation point, a 250 ms blank, and the second item in the pair. After presentation of the second item, the ratings scale appeared on the screen, and the next trial was initiated as soon as a response was given. Items were displayed for one and a half seconds. Each participant rated a fourth of the possible pairs of items. The set of pairs was presented twice in different random orders, creating about 700 trials for each participant in this part of the study.

Results

On the classification test, participants correctly classified 56% of test items as grammatical or ungrammatical (MSE = 1.54%). This level of performance is similar to previous studies, and is significantly better than chance, $t(15) = 35.9$, $p < .0005$.

Similarity ratings were averaged and transformed to form a single dissimilarity matrix. Multidimensional scaling was used to derive a spatial representation of the stimuli which preserved the basic structure of the dissimilarity matrix. The optimal spatial configuration involved a 3D space with a Euclidean distance metric ($r = 2$) (cf. Shepard, Romney, & Nerlove, 1972). The stress of this configuration was 0.18, representing a reasonable fit (zero stress represents a perfect fit; cf. Kruskal, 1964; Krzanowski, 1993).

We used non-linear regression to fit items in the psychological space to participants' grammaticality endorsements, using the GCM model described by Equations (1–3) (cf. Nosofsky, 1989, 1992). Since AGL tasks, including the present one, involve training with grammatical items, but not with ungrammatical items, Equation (1) represents a reduced version of the GCM which compares each item to old items from only a single category.

$$P(R_1 | S_i) = b \sum_{j \in C_1} s_{ij} \tag{1}$$

$$s_{ij} = e^{-c \cdot d_{ij}^p} \tag{2}$$

$$d_{ij} = \left[\sum_{m=1}^{\dim} w_m |x_{im} - x_{jm}|^r \right]^{1/r} \tag{3}$$

According to this model, the probability that a test item will be classified as grammatical is determined by adding up the similarity of the test item to each training item. Similarities between items, s_{ij}, are computed from the distances, d_{ij}, between items in psychological space. p distinguishes between possible forms of the function relating similarities to istances (usually Gaussian or exponential). The free parameters of the model are b, c , and the dimension weights, w_m(which are constrained to sum to 1). The best fit was obtained with an exponential function relating distance to similarity (i.e. $p = 1$), with an R^2 value of 0.42. This is a measure of the variance accounted for by a similarity-only model, with no provision for additional rule-based knowledge.

The extent to which the GCM fits grammaticality endorsements reveals how well similarity effects capture classification performance. Tests for similarity and rule-based knowledge proceeded in two stages. The first stage followed the analysis of Knowlton and Squire (1996), except that where Knowlton and Squire measured similarity by counting the number of bigrams and trigrams shared between test and training items, we used GCM-derived values of similarity. Grammatical and ungrammatical test items were split into high- and low-similarity subsets, cutting at the median similarity values of each group. Mean similarity values for the high- and low-similarity groups were significantly different, for both grammatical and for ungrammatical items, $t(14) > 5.5$, $p < .0005$ for both. There was no significant difference in similarity values for grammatical and ungrammatical items within either the high similarity or the low similarity groups, $t(14) < 1.9$, $p > .05$ for both. A two-way ANOVA on grammaticality endorsement rates showed main effects of both grammaticality, $F(1,15) = 14.88$, $p = .002$, and similarity, $F(1,15) = 30.13$, $p < .0005$. The interaction term just missed significance, $F(1,15) = 3.65$, $p = .076$. These results show that participants were more likely to endorse an item as grammatical if it was more similar to training items, but also if it actually complied to the rules of the grammar used. These findings are in accord with existing experimental work (Knowlton & Squire, 1996; Brooks & Vokey, 1991), but they do not necessarily mean both rule-based and exemplar-based learning is taking place. Despite Knowlton and Squire's efforts to design stimuli in which grammaticality and similarity are dissociated, there remains a small correlation between similarity and grammaticality. This correlation confounds the interpretation of the ANOVA results, even though the correlation between similarity and grammaticality is non-significant, 0.25, $p > .15$. What we really want to know is, given the GCM model of similarity, is there any evidence that participants also employed some other knowledge when making grammaticality judgements?

To test whether there were effects of grammaticality that could not be accounted for by similarity and vice versa, we ran several regression analyses with grammaticality endorsements as the dependent variable. Similarity was modelled by the GCM model above, which included four parameters (*b*, *c*, and two dimension weights). Rule-based knowledge of grammaticality was modelled with a simple additive term indicating whether each item really was or was not grammatical. Regressions minimised R^2. Comparisons between models are based on F-change statistics (Howell, 1996).

Table 1 presents statistics for regression models involving similarity, grammaticality, or both, and also presents statistics for comparisons between nested models. These comparisons test whether the addition of a factor makes a significant improvement to a simpler model. The main finding is that grammaticality contributes nothing over and above the contribution of similarity, but the reverse is not true. The simplest interpretation is that grammaticality judgements in this study were based on the similarity of test items to training items, and that participants extracted no additional knowledge of the abstract grammar underlying the training stimuli. The failure to confirm an independent contribution of grammaticality in the regression analysis highlights a shortcoming of the more usual ANOVA analysis, which fails to unconfound factors which are partially correlated (cf. Bogartz, Shinskey, & Speaker, 1997).

Experiment 2

Participants

Sixteen University of Oxford students took part in the study.

Materials

The artificial grammar was identical to the one used in Experiment 1, but here the training and test strings were mapped to angles between connected lines to produce stimuli like that shown in Figure 3. The letters in Knowlton and Squire's material, VXJT, were identified with the angles

Table 1: Regression Analysis for Variables Predicting Grammaticality Endorsements in Experiment 1.

Model(s)	R^2	*df*	R^2-*Ch*	*df-Ch*	*F-Ch*	*p-Ch*
Sim Only	0.42	4	0.42	3	6.25	.002*
Gram Only	0.10	2	0.10	1	3.05	.09
Sim + Gram	0.44	5	0.02	1	0.79	.38
Gram + Sim	0.44	5	0.34	3	5.01	.01 *

Ch statistics for simple models reflect improvement over the null model with a single parameter for mean endorsement rate. *Ch* statistics for combined models reflect the improvement achieved by adding a second factor to a simpler model including only the first factor.

* A significant *p-ch* value means the more complex model is significantly better than the simpler model.

75°, 150°, 225°, and 300°, respectively. Unlike previous AGL studies, the visual parts of these stimulus items were context-dependent, in that the angle of each line segment was determined relative to the angle of the previous segment.

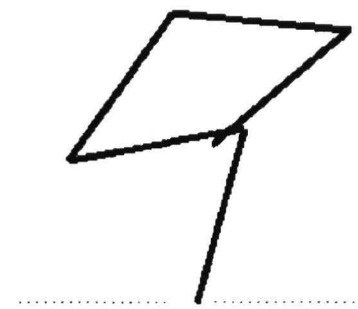

Figure 3: Sample stimulus for Experiment 2, corresponding to the string VXJTJ.

Procedure

The procedure was identical to that of Experiment 1 except that the instructions were modified as appropriate for the different stimuli. Also, there was no central fixation point between stimuli.

Results

Analysis proceeded as in Experiment 1. On the classification test, participants correctly classified 55.29% of test items as grammatical or ungrammatical (MSE = 1.57%), which is significantly better than chance performance, $t(15) = 35.2$, $p < .0005$. The level of performance on Experiment 2 was not different from that for Experiment 1, $t(30) = 0.3$, $p > .75$.

The best fit for the GCM was obtained with a city block distance metric and an exponential function relating distance to similarity. This combination of parameters produced an MDS stress value of 0.20 and a GCM R^2 value of 0.55.

Again, grammatical and ungrammatical test items were split at the median GCM similarity values within each group. Mean similarity values for the high- and low-similarity groups were significantly different, for both grammatical and for ungrammatical items, $t(14) > 4.5$, $p < .0005$ for both. There was no significant difference between grammatical and ungrammatical items within either the high similarity or the low similarity group, $t(14) < 0.9$, $p > .35$ for both. A two-way ANOVA on grammaticality endorsement rates showed main effects of both grammaticality, $F(1,15) = 11.5$, $p < .005$, and similarity, $F(1,15) = 104$, $p < .0005$. There was no significant interaction, $F(1,15) = 1.45$, $p = .248$. As in Experiment 1, participants were more likely to endorse an item as grammatical if it was more similar to training items, but also if it actually complied to the rules of the grammar used.

Regression models were fit to grammaticality endorsements rates to test whether there were effects of grammaticality that could not be accounted for by similarity and vice versa. Regression results are summarised in Table 2. In this study, unlike Experiment 1, similarity on its own does not account for grammaticality judgements. The best model includes both similarity and grammaticality, and removal of either factor results in a significantly worse model. The significance of both factors suggests that grammaticality endorsements were based on a combination of similarity to training items and some other rule-based knowledge extracted from the training items. Alternatively, it is always possible that grammaticality judgements were based on some other (unknown) single knowledge representation which is partially correlated with both GCM similarity and with actual grammaticality, but this hypothesis can be evaluated only with respect to a particular theory of what that knowledge representation might be.

Discussion

Classification of novel items in two studies was largely predictable on the basis of similarity between test and training items. We employed an independently-motivated exemplar model of categorisation in conjunction with empirical similarity data, avoiding criticisms levelled at ad hoc models of similarity. In Experiment 2, but not in Experiment 1, similarity alone was not sufficient to account for generalisation performance. This suggests either that summary knowledge (e.g. rules) plays a role in some generalisation tasks but not in others. The finding of *independent* effects of both similarity and grammaticality in Experiment 2, in an analysis which statistically unconfounds the two factors, confirms a rule-based component of knowledge underlying generalisation in at least some AGL tasks. Our two studies were identical except for the visual form of the stimuli, suggesting that the relative contributions of similarity and rule-based knowledge can vary with specific stimulus attributes.

The failure in Experiment 1 to find independent effects of grammaticality over and above the effect accounted for by similarity highlights a shortcoming of ANOVA. Conclusions of previous studies finding main effects of both similarity and grammaticality are confounded by non-zero correlations between the two factors. This confound was avoided in our analysis by comparing nested regression models in order to test effects of one factor over and above effects accounted for by the other factor.

Table 2: Regression Analysis for Variables Predicting Grammaticality Endorsements in Experiment 2.

Model(s)	R^2	*df*	R^2-*Ch*	*df-Ch*	*F-Ch*	*p-Ch*
Sim Only	0.59	4	0.59	3	12.66	.000*
Gram Only	0.08	2	0.08	1	2.40	.13
Sim + Gram	0.70	5	0.10	1	8.42	.008*
Gram + Sim	0.70	5	0.62	3	16.91	.000*

See legend for Table 1.

Acknowledgements

Todd Bailey was supported by a grant from the McDonnell-Pew Centre for Cognitive Neuroscience. Emmanuel Pothos was supported by the UK Medical Research Council (reference number: G78/ 4804), the Bodossaki foundation, and the A. S. Onasis foundation (reference: Group S-076/1996-97). We wish to thank Nick Chater, David Firth, Kim Plunkett, and Martin Redington for their assistance.

References

Ashby, G. F. & Alfonso-Reese, A. L. (1995). Categorisation as Probability Density Estimation. *Journal of Mathematical Psychology, 39,* 216-233.

Bogartz, R. S., Shinskey, J. L., & Speaker, C. J. (1997). Interpreting infant looking: The event set X event set design. *Developmental Psychology, 33,* 408-422.

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

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General, 17,* 316-323.

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.

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.

Goodman, N. (1954). *Fact, fiction and forecast.* London: Athlone Press.

Hahn, U. (1996). *Cases and Rules in Categorisation.* DPhil thesis, Department of Experimental Psychology, University of Oxford.

Hahn, U. & Chater, N. (1998). Similarity and Rules: Distinct? Exhaustive? Empirically Distinguishable?. *Cognition, 65,* 197-230.

Howell, D. C. (1996). *Statistical Methods for Psychology.* Belmont, CA: Duxbury Press.

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.

Kruskal, J. B. (1964). Multidimensional scaling by optimising a goodness-of-fit to a nonmetric hypothesis. *Psychometrika, 29,* 1-27.

Krzanowski, W. J. (1993). *Principles of Multivariate Analysis: A User's Perspective.* Oxford: Clarendon Press.

McKinley, S. C. & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 128-148.

Medin, D. L. (1986). Comment on "Memory on memory storage and retrieval processes in category learning." *Journal of Experimental Psychology: General, 115,* 373-381.

Medin, D. L. & Schaffer, M. M. (1978). Context Theory of Classification Learning. *Psychological Review, 85,* 207-238.

Nosofsky, R. M. (1988a). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, Cognition, 14,* 700-708.

Nosofsky, R. M. (1988b). Similarity, Frequency, and Category Representation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14,* 54-65.

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.

Nosofsky, R. M. (1990). Exemplars, prototypes, and similarity rules, Cognitive Science, Indiana University Research Report 17.

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

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43,* 25-53.

Perruchet, P. (1994). Defining the knowledge units of a synthetic language: Commentary on Vokey and Brooks (1992). *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20,* 223-228.

Perruchet, P. Gallego, J., & Pacteau, C. (1992). A reinterpretation of some earlier results for abstractiveness of implicitly acquired knowledge. *Quarterly Journal of Experimental Psychology, 44A,* 193-210.

Perruchet, P., Pacteau, C. (1990). Synthetic Grammar Learning: Implicit Rule Abstraction or Explicit Fragmentary Knowledge?. *Journal of Experimental Psychology: General, 119,* 264-275.

Pothos, E. M. & Chater, N. (1997). Generality of the Abstraction Mechanisms in Artificial Grammar Learning. Submitted manuscript.

Reber, A. S. (1967). Implicit Learning of Artificial Grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855-863.

Redington, F. M. (1996). *What is Learnt in Artificial Grammar Learning?* DPhil thesis, Department of Experimental Psychology, University of Oxford.

Redington, F. M. (1997). No evidence for the acquisition of abstract (surface-independent) rule-based knowledge: Commentary on Knowlton and Squire (1996). Submitted manuscript.

Servan-Schreiber, E. (1991). *The Competitive Chunking Theory: Models of Perception, Learning, and Memory.* Doctoral Dissertation, Department of Psychology, Carnegie-Mellon University.

Servan-Schreiber, E., Anderson, J. R. (1990). Learning Artificial Grammars With Competitive Chunking. *Journal of Experimental Psychology: Learning Memory and Cognition, 16,* 592-608.

Shepard, R. N. (1980). Multidimensional Scaling, Tree-Fitting, and Clustering. *Science, 210:* 390-398.

Shepard, R. N. (1987). Toward a Universal Law of Generalisation for Psychological Science. *Science, 237,* 1317-1323.

Shepard, R. N., Romney, A. K., & Nerlove, S. B. (1972). *Multidimensional Scaling, Volume I: Theory.* New York: Seminar Press.

Shin, H. J. & Nosofsky, R. M. (1992). Similarity-Scaling Studies of Dot-Pattern Classification and Recognition. *Journal of Experimental Psychology: General, 121,* 278-304.

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

Vokey, J. R., & Brooks, L. R. (1994). Fragmentary knowledge and the processing specific control of structural sensitivity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 1504-1510.

Watanabe, S (1985). *Pattern recognition: human and mechanical.* New York: Wiley.