UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Generality of the Abstraction Mechanisms in Artificial Grammar Learning

Permalink
https://escholarship.org/uc/item/065477ds

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 20(0)

Authors
Pothos, Emmanuel M.
Chater, Nick

Publication Date
1998

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Generality of the Abstraction Mechanisms in Artificial Grammar Learning

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

Nick Chater (nick.chater@warwick.ac.uk)
University of Warwick; Department of Psychology
Coventry CV4 7AL, UK

Abstract

Artificial grammar learning (AGL; Reber, 1989) has been a major experimental paradigm for the study of human induction processes. In this work we investigate the extent to which the learning mechanisms involved in AGL are general, an issue important to the ecological validity of AGL research. We have used three kinds of stimuli: Letter strings (the standard in AGL work), city sequences that corresponded to routes of an airline company, and shapes that were presented so that later shapes in a sequence contained all previous ones. We compared overall accuracy and patterns of error in these domains to find that performance was not different. The implications of this finding for existing theories of AGL and proposed relations to other cognitive mechanisms are discussed.

Introduction

Central to the existence of all living creatures is the ability to abstract the essential structure in a domain of instances in order to successfully generalize to other new instances. That living agents survive implies that nature has somehow solved the problem of induction. The abundance of successful inductive learners makes it even more unfortunate that our understanding of induction is still inadequate. For instance, Pinker notes: (1979; p.278) "Induction has been called 'scandalous' because any finite set of observations supports an intractably large number of generalizations."

As early as the late fifties, Miller (1958) presented participants with a set of letter strings generated by a finite state language (Chomsky & Miller, 1958) to find that learning was more efficient when compared with learning of random strings. This was, presumably, by virtue of the fact that "...the sequences formed according to elaborate and restrictive rules were more redundant, carried less information per syllable, than the random syllables." Reber (1967) not only replicated this finding but also extended it in the case where participants had to discriminate between strings that complied with the rules of the language and strings that violated these rules in certain ways. The training of Reber's participants involved presenting a subset of the possible strings that could be generated with the grammar. These studies initiated a long tradition of AGL experiments. The AGL paradigm promised to provide a more principled way of understanding human abstraction processes, because the domains used (strings generated from finite state languages) could be very precisely described in information theory terms (e.g., Miller, 1958).

Several different theories have been proposed to account for AGL competence. The original claim by Reber and his colleagues (see Reber, 1989, for a review) has been that participants learn in training something of the abstract, rule structure of the finite state language used to create the stimuli. This view was corroborated by "transfer" experiments, where the symbols used in training were different from the symbols used in test. Dulany, Carlson, & Dewey (1984) have instead argued that participants acquired "correlated grammars", that is a set of "microrules" which generally approximated the true grammar, but might at the same time include unrepresentative or even wrong rules. A similar approach from a different point of view has been offered by Perruchet and Pacteau (1990) who suggested that all participants learn in the training part is simply information about which bigram fragments are allowed.

Brooks and Vokey (1991) suggested that grammaticality decisions in test do not so much depend on whether an item is grammatical or not-grammatical, but rather on whether it is more or less *similar* to the training items. They have found that both similarity and grammaticality are important, and these results have been replicated in later research (Knowlton & Squire, 1996; Pothos & Bailey, 1997).

The theoretical controversy as to which account best reflects human performance is still going on; however, all theories mentioned are specified independently of the particular instantiation of the stings of a finite state language. Thus, a direct prediction of existing AGL theories is that performance should not change with different stimulus domains, as long as the grammar used is the same.

We used the Reber and Allen (1978) grammar to generate three sets of stimuli, letter strings (as in the Reber and Allen, 1978, study), city sequences that corresponded to routes of an airline company (referred to as "cities" stimuli below) and shapes that were presented so that later shapes in a sequence contained all previous ones (referred to as "shapes"). Our choice for this grammar was motivated partly because this is a grammar that has been widely used (e.g. Dulany et at., 1984; Perruchet & Pacteau, 1990) but also because it has been designed in a way that allows an investigation into the particular pattern of errors.

There were some a priori reasons why performance might be specific to the type of stimuli. For instance, it has been suggested that AGL is essentially equivalent to priming, rather than reflecting a more profound type of learning mechanism (Knowlton & Squire, 1996). Roughly speaking, implicit memory, or priming, of previous events is assessed by looking at effects of these events on a later task, in the absence of any explicit recall of the events. Object decision priming has been suggested to be mediated by a "perceptual representation system" (PRS) that essentially describes the structural dependencies in an object (Schacter et al., 1991), and so that, for instance, no priming for "impossible" objects can be observed (since the PRS cannot describe the global structure of an impossible object). Thus, if AGL does indeed reflect priming, performance with the shapes stimuli would be expected to be superior, since the regularities with these stimuli have a more salient spatial structure, which would be better "understood" by the PRS. Our "shapes" stimuli consisted of nested elements precisely to stress the impression of a coherent whole, and thus emphasize spatial dependencies. Conversely, in a transfer condition, if the priming view of AGL is accurate, performance would be particularly bad, relative to performance with stimuli where the sequential symbolic structure of the strings was more obvious (as is the case with letter strings, for example).

Also, with the cities stimuli one could anticipate that encoding of the regularities due to the artificial grammar we used would be strongly affected by background knowledge biases, relating to the relative plausibility of different routes. Since we did not construct the stimuli so that the routes might reflect any realistic constraint, trying to interpret the task pragmatically would be thus likely to reduce overall accuracy; such biases would also be expected to be more potent in a transfer condition since direct information about the items is compromised by the change in the city names used.

Previous investigations on the generality of the learning processes involved in AGL have mostly focused on identifying some of the conditions in which AGL-type learning was possible. For instance, Altmann, Dienes and Goode (1995) used musical tones and graphic symbols[1] in some of their conditions. Altmann et al. reported that participants could successfully generalize knowledge acquired from musical tones to letter sequences and vice versa, or from sequences of nonsense words presented auditorially to sequences of graphic symbols (see also Whittlesea and Wright, 1997, for a more recent similar investigation). In the work reported below we were interested in comparing directly performance with different types of stimuli, to address specifically the question of how general are the abstraction mechanisms involved.

Method

Design

Predictions for differences in the learning mechanisms apply both to the ordinary AGL paradigm, where there is no change in the symbols used from training to test, and the transfer paradigm, where the symbols are changed. We thus report experiments employing both paradigms. In each case, a 3 x 8 mixed design was used, with training and testing on the three sets of stimuli as the between-subjects factor and performance on eight different subsets of the test strings as a within subjects factor. Our interest was in comparing performance with different types of stimuli, rather than assessing learning of the grammar, therefore we did not use controls (see Redington & Chater, 1996, for an extensive discussion of methodological issues in AGL).

Materials

In all conditions we created the stimuli according to the Reber & Allen (1978) grammar[2], shown in Figure 1. In the condition where the there was no change in vocabulary from training to test, the letters used were RXMSV and the "shapes" and "cities" stimuli were constructed by mapping letters to shapes and cities respectively. So, RXMSV corresponded to a circle, a hexagon, a concave shape, a square and a rhombus in the one case, and in the other to Athens, Berlin, London, Paris and Madrid. The shapes were chosen so that they were of roughly similar salience and, likewise, the cities were intended to be well known European capitals and representative of most of Europe. Examples of stimuli in the different conditions are shown in Figure 2.

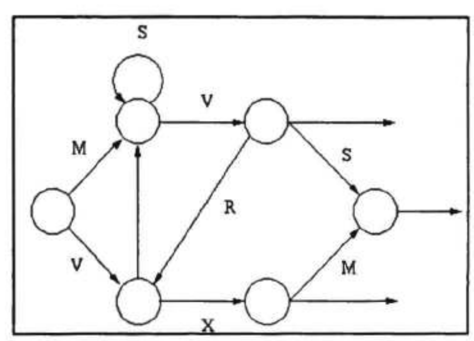

Figure 1: The finite state grammar used in the Experiments.

[1] But the symbols in the graphical stimuli were presented sequentially as in letter strings, thus remaining perceptually distinct.

[2] Details for the motivation for this grammar and the ways test items were constructed are given in Reber and Allen (1978).

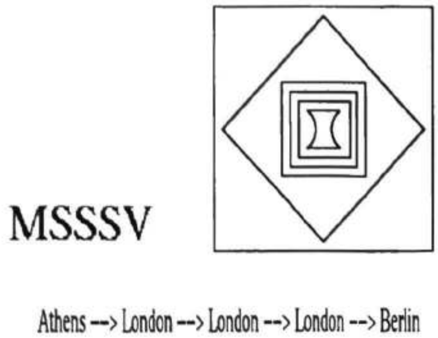

Figure 2: Examples of the types of stimuli used.

In the transfer condition, the symbols used to create the stimuli were changed from training to test. In selecting suitable new shapes, the same considerations as in the no-transfer condition applied, while in the "cities" experiments, we used a set of American cities in training, and preserved the European ones in test.

Participants

Ten participants were tested with each type of stimuli in the no-transfer condition, and twice that number in the transfer one. Participants were members of the University of Oxford who either volunteered to participate or were paid 1.50 pounds traveling expenses. They were all naive regarding the experimental procedure and the particular materials used. The experiment lasted for approximately half an hour, including a debriefing session, and all participants were tested individually.

Procedure

We attempted to replicate the original procedure of Reber & Allen (1978), especially with respect to the format of presentation. In the no-transfer condition, in the first part of the experiment, participants were presented with a folder containing a set of stimuli and instructions. The instructions emphasized that they had to "pay the utmost attention to the letter strings/routes/ shapes" (depending on the particular set of stimuli used) but nothing more; also, although they were told that there would be a second part to the experiment, no information was given as to what would be required of them. In the case of the "cities" stimuli, participants were told that "city names correspond to where a plane of the company was at noon, on successive days." The use of such vague instructions was intentional, in line with the consistent finding in the literature that "The circumstances under which subjects are most clearly sensitive to general principles of the grammar are those that require subjects to do the least with stimuli and that therefore favor incomplete encoding of particular items." (Whittlesea & Dorken, 1993). While reading the instructions, they were encouraged to ask questions if anything was not clear. The training set consisted of 20 unique grammatical items presented three times in a random order. Once the instructions had been read, participants turned to the first sheet in the folder, which they saw for approximately 10 seconds. The experimented tapped on the table they were working on, to indicate that they had to proceed to the next item. On average, about 10 minutes were required for the training phase of the experiment.

Participants then received another folder and more printed instructions. The folder contained 50 unique items presented twice in a random order, half of which were grammatical and the other half violated the grammar in specific ways (for details see Reber and Allen, 1978). It was emphasized that the order in which cities (/letters/shapes) in each of the presented items was determined by a rather complex set of rules, that half of the items in the second folder complied with these rules while the other half violated them and that they were to discriminate 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 pragmatic stimuli, it was added that the rules were such so as 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 so as to encourage participants not to bias their decisions by the relative frequency of specific cities across items. Participants indicated their grammaticality decisions by attaching a post-it note with the letters G for "good" and N for "not good" to the corresponding items. There was no time limit in making any one decision.

The main procedural difference in the transfer condition was that the experiment was then computer-based so that participants saw the stimuli on a computer screen, and their answers were indicated by pressing labeled buttons. The training instructions were identical, while in test it was simply added that the new items would involve different symbols. Also, with the "cities" stimuli, to justify the mapping from American to European cities, a story was used where a European airline company decided to use the same rules for constructing routes as the American company, the routes of which were presented in training.

Results

In the no transfer condition, performance was significantly above chance with all types of material (one sample two-tailed t-tests with 9 degrees of freedom resulted in t values significant at the 0.0005 level). Nonetheless, a one way ANOVA with "type of stimuli used" (that is letters, cities or shapes) as a between subjects factor was not significant ($F(2,27) = 0.698$, $p > .5$, $MSE = 0.006$). Figure 3 illustrates this result. The differences in overall accuracy were also investigated by power analyses, whereby for all pairs of stimulus types an estimate of the sample size that would lead to $p=.01$, power=0.95 difference, was calculated. In all cases, the required sample size for such a difference was no less than about 150 subjects, thus providing additional support for a lack of observed differences.

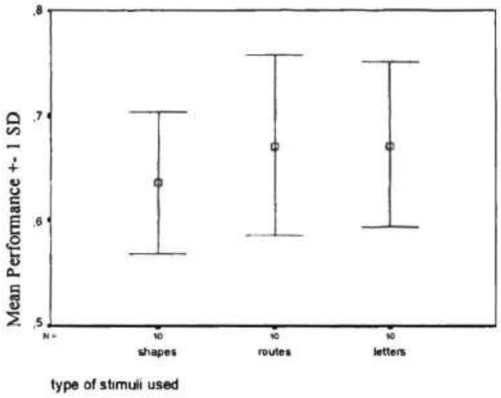

Figure 3: Performance with different types of stimuli.

The Reber and Allen (1978) stimuli that we used were designed so that non-grammatical items were constructed by violating grammatical items in specific ways. Therefore, the items used in the test part of the experiment could be split in the following mutually exclusive subsets: Items with violations in the first position ("vio first, " in Figure 4), violations in the second position ("vio second"), deep violations, violations in the next to last position ("vio next to last"), violations in the last position ("vio last") and items that were spelt backwards ("backwards"). The grammatical items used in the test part were distinguished depending on whether they had been presented in the training part or not ("seen" and "rest of G items").

Performance in the different subsets for the three types of stimuli used is shown in Figure 4. Repeated measures ANOVA with "performance in different subsets of test items" as a within subjects factor were conducted for each type of stimuli separately. In all three models the homogeneity of variance assumption could not be maintained (as assessed by the Mauchly Sphericity Test) thus we had to resort to a more conservative test. For the "letters" and the "cities" stimuli the F-ratio was not significant by the Greenhouse-Geisser test. (for "letters": $F(3.02, 63) = 2.24$, $p=.105$ and for "cities": $F(2.66, 63) = 1.14$, $p=.349$). In the case of the "shapes" stimuli very low performance on items with violations in the next to last position led to a highly significant difference ($F(2.63, 63) = 7.69$, $p=.001$). We then looked at whether there was an interaction between type of stimuli used and performance in different subsets of the test items. A mixed factorial design was run with "type of stimuli used" as the between subjects factor and "performance for different subsets of test items" as a within subjects factor. Again, the Mauchly Sphericity Test was significant ($W < 0.05$) so we had to use a test with fewer degrees of freedom. However, the interaction between "type of stimuli used" and "performance for different subsets of test items" was not significant ($F(7.53, 189) = 1.69$, $p>.1$) by the Greenhouse-Geisser test.

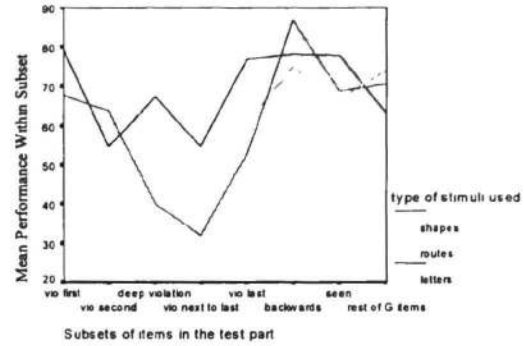

Figure 4: Performance on different subsets of the test set.

The overall pattern of results for the transfer conditions was similar, albeit distinguished by a decrease in accuracy, as is typically found with transfer experiments (Whittlesea & Dorken, 1993; Altmann et al., 1995). Overall performance was in general only slightly above chance (Figure 5) and one-sample t-tailed tests against chance were significant in the cases of the cities and letters stimuli, but not in the case of shapes (cities stimuli: $t(19) = 3.831$, $p=.001$; letters stimuli: $t(19) = 2.57$, $p=.019$; shapes: $t(19) = 0.711$, $p>.7$). Despite the fact that it might look as if learning is possible with stimuli sets such as cities and letters and not possible with the shapes, as before, the differences in performance across conditions were not significant ($F(2,57) = 2.537$, $p =.09$, $MSE = 0.003$). Power analyses again revealed that at least 150 participants would have been required to provide a $p=.01$, power=0.95 difference for a given comparison between performance on different kinds of stimuli (largest sample estimate was more than 10000).

Likewise, a mixed-factorial design ANOVA with "type of stimuli used" as the between participants factor and "performance for different subsets of test items" as a within participants factor, was not significant ($F(14, 399) = 1.43$, $p =.138$), thus replicating the essential findings of the no-transfer condition (see Figure 6).

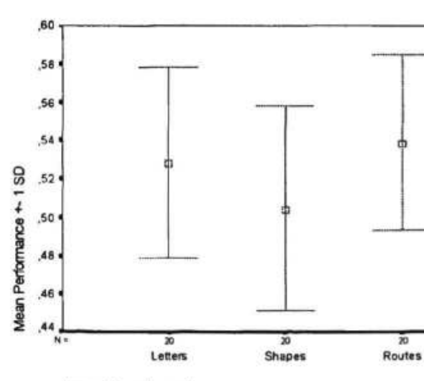

Figure 5: Mean performance with each type of stimuli, in the transfer condition.

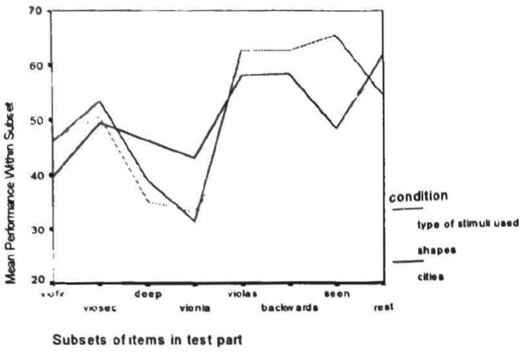

Figure 6: Performance on different subsets of the stimuli for the transfer condition.

Discussion

We compared an overall accuracy measure for the different kinds of stimuli and also we looked at whether the patterns of error were any different across stimulus domains, both in a no-transfer experimental paradigm and a transfer one. Differences were expected by taking note of the possible connection of AGL with priming (Schacter et al, 1991) and also the fact that a realistic content was likely to introduce general knowledge biases that were irrelevant with the actual regularities in the stimuli.

It was concluded that performance did not seem to depend on the type of material used, the way items were constructed or the existence of any possible biases. This result is significant for existing theories of AGL since they all assume that the learning processes involved in successful generalization from training to test depend only on the human processor being able to access the symbolic structure of the stimuli. Also, it supports the use of AGL as an experimental paradigm for the study of inductive inference; the finding that AGL does not depend on differences with the types of stimuli investigated suggests that AGL-type of learning might well be directly generalizable to more realistic learning situations as well.

Future work will further address the conditions that might be affecting AGL. For instance, the stimuli in this study were all constructed so as not to affect in any way the symbolic salience of the stimuli, in line with the observation that all existing theories of AGL depend on such information being available. Thus, an interesting question that follows is whether learning would be compromised in situations where such information is less available.

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.

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 DPhil thesis, University of Oxford.

Chomsky, N. & Miller, G. A. (1958). Finite State Languages. *Information and Control, 1,* 91-112.

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.

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.

Miller, G. A. (1958). Free Recall of Redundant Strings of Letters. *Journal of Experimental Psychology, 56,* 485-491.

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

Pinker, S. (1979). Formal models of language learning. *Cognition, 7,* 217-283.

Pothos, E. M. & Bailey, T. M. (1997). Rules vs. Similarity in Artificial Grammar Learning. In *Proceedings of the Similarity and Categorisation Workshop 97,* 197-203, University of Edinburgh.

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.

Redington, M. & Chater, N. (1996). Transfer in Artificial Grammar Learning, Methodological Issues and Theoretical Implications. *Journal of Experimental Psychology, General, 125,* 123-138.

Schacter, D. L., Cooper, L. A., Delaney, S. M., Peterson, M. A., & Tharan, M. (1991). Implicit memory for possible and impossible objects, Constraints on the construction of structural descriptions. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 17,* 3-19.

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.

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.