Learning & Behavior
2003, 31 (1), 98-106

"Artificial grammar learning" in pigeons: A preliminary analysis

WALTER T. HERBRANSON
Whitman College, Walla Walla, Washington

and

CHARLES P. SHIMP
University of Utah, Salt Lake City, Utah

An avian analogue to human *artificial* or *synthetic grammar learning* (Reber, 1967) was developed. Pigeons viewed horizontal strings of three to eight colored letters. These strings either conformed to Reber's artificial grammar or violated it in one or two locations. Pigeons categorized the letter strings as *grammatical* (left keypeck) or *nongrammatical* (right keypeck). Overall accuracy of categorization was above chance to both familiar training strings and to novel transfer strings, thereby satisfying a conventional criterion for learning an *abstract concept.* The results support a *multiple mechanisms* point of view according to which pigeons, like humans, learn both abstract concepts and specific strings, or specific parts of strings, in artificial grammar learning tasks.

A common belief about the adaptive function of a general category or concept is that it reduces the amount of information processing in which a human would otherwise have to engage if he or she had to process the full uniqueness of each stimulus he or she encountered (Smith & Jonides, 2000). It would seem that nonhuman animals would similarly benefit from such an adaptive mechanism, and indeed, it now seems clear that some nonhuman animals can learn general concepts or categories. Pigeons, for example, can learn several kinds of concepts, including same/different (Cook, Cavoto, & Cavoto, 1995; Cook, Katz, & Cavoto, 1997), matching to sample (Wasserman, Hugart, & Kirkpatrick-Steger, 1995; Wright, 1997, 2001), the difference between local and global levels of visual attention (Fremouw, Herbranson, & Shimp, 1998, 2002; Shimp, Herbranson, & Fremouw, 2001), naturalistic visual concepts (Herrnstein & Loveland, 1964; Herrnstein, Loveland, & Cable, 1976), and general decision rules by which to approximate optimal solutions to multidimensional ill-defined categorization tasks (Herbranson, Fremouw, & Shimp, 1999, 2002).

All these demonstrations that pigeons can learn concepts were based on a two-part criterion. First, pigeons learned training exemplars to a better-than-chance degree, and second, they transferred that performance to novel exemplars to a better-than-chance degree. This two- part criterion has the utilitarian virtue of capturing at least part of what, in plain English, it means to use a concept: to place diverse nonarbitrary exemplars, both familiar and novel, into a single category. It should be acknowledged that the conventional two-part operational definition of a concept does *not* address many deeper theoretical questions about the nature of the mechanisms underlying the use of a concept. Thus, much remains to be learned about how a naturalistic category is learned or, indeed, even about what is learned. Despite these acknowledged limitations, the two-part definition provides the classic advantages of an operational meaning of a concept, and it is more than adequate for our present purpose, which is merely to demonstrate a new avian analogue of a familiar experimental paradigm in the human cognition literature.

This paradigm is variously referred to as involving an *artificial grammar*, a *miniature grammar*, a *finite state grammar*, or a *synthetic grammar*. It is important at the outset to address a terminological issue that inevitably arises owing to the conventional use in the human literature of a term like *artificial grammar*. Chomsky (1957) and Chomsky and Miller (1958), who appear to have introduced the concept, seem to have hoped initially that the seemingly preternaturally complex grammar of natural language could be modeled by artificial grammars that involve simple rules by which symbol strings can be generated. They seem quickly to have decided, however, that the one is not just a simpler version of the other (Crowther-Keyck, 1999). In particular, the meaning of meaning is different in the two

---

This research was supported in part by a grant from the University of Utah Research Committee and by an NIMH predoctoral fellowship. This work was based on part of the first author's doctoral dissertation (Herbranson, 2000). The authors thank Thane Fremouw for stimulating conversations on issues related to this research and for his help on a variety of practical matters in the laboratory during early stages of this research.

Correspondence concerning this article should be addressed to C. P. Shimp, Department of Psychology, 380 S. 1530 E., Rm 502, University of Utah, Salt Lake City, UT 84112-0251 (e-mail: charlie.shimp@psych. utah.edu).

—Accepted by the previous editorial team of Ralph R. Miller.

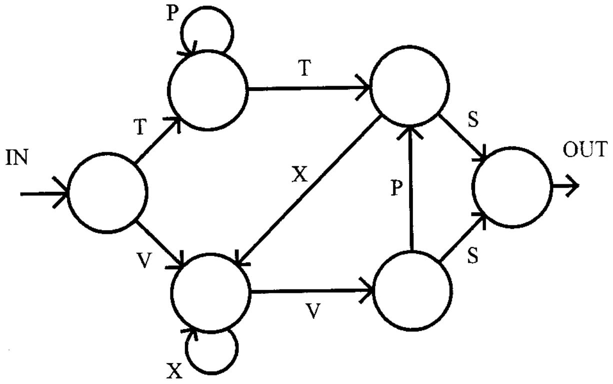

Figure 1. The artificial grammar that generated *grammatical* character strings of lengths 3–8. Adapted from Chomsky and Miller (1958) and Reber (1967).

cases, with meaning in the artificial case limited to how different categories of strings of arbitrary characters relate differently to reinforcing consequences, such as, for a human participant, social approval by an experimenter.[1]

Reber (1967) pioneered the use of artificial grammars in experimental settings. He successively presented many, but not all, strings that could be generated by an artificial grammar to college undergraduates and instructed them to memorize those strings but did not inform them about the existence of a grammar from which the strings were derived. Figure 1 shows the specific artificial grammar Reber (1967) used (see also Chomsky & Miller, 1958, p. 102). This grammar is perhaps the modal grammar used thereafter in research on artificial grammars with human participants. The grammar has the Markovian property that the state y_{n+1} on trial $n + 1$ depends only on state x_n on trial n. Grammatical strings are generated by starting at the *in* arrow and moving through the grammar along the arrows, with each transition adding a letter to the string, until the process exits via the *out* arrow.

After having observed training strings a sufficient number of times to be able to memorize them, participants were told that the training exemplars had been generated by a system of rules and were asked to categorize novel exemplars as conforming to, or as not conforming to, that same set of rules, which were still not described. Thus, the participants were shown novel, not previously viewed, exemplars of the same grammar. Intermingled among these novel grammatical exemplars were other novel strings that violated the artificial grammar in one or two locations.

Reber (1967) found that the participants, on the average, correctly categorized novel exemplars as grammatical or nongrammatical about 79% of the time (Experiment 2, p. 861), significantly above the chance value of 50%, but far below a perfect value of 100%. Overall accuracy of categorization in experiments of this general type varies over a considerable range, depending on specific procedural details. For example, the average percentage for categorization of strings in McAndrews and Moscovitch (1985) was only 57%. For this reason, the term *artificial grammar learning* is somewhat of a misnomer. Few if any researchers using this method now seem to believe that human participants learn an artificial grammar in its entirety, although a significant number do seem to believe that humans learn various bits and pieces of artificial grammars.

Human participants often fail to describe verbally the artificial grammar when asked to describe the set of rules that generated the exemplars they observed. They may even fail verbally to summarize accurately their own rules, which may be inferred from their categorizations (Dienes, Altmann, Kwan, & Goode, 1995; but see Dulany, Carlson, & Dewey, 1984). Reber (1967) concluded that participants learn rules of the artificial grammar, without knowing how they do so (Reber & Lewis, 1977). For this reason, Reber (1967) argued that artificial grammar learning more closely resembles perceptual learning than associative learning. He views artificial grammar learning as an unaware, implicit acquisition of abstract rules and, at least initially, specifically rejected the idea that participants learn only specific responses to specific stimuli or specific components of stimuli.

Others have suggested instead that artificial grammar learning may also involve, at least in part, learning specific features of grammatical character strings (Knowlton & Squire, 1994; Meulemans & Van der Linden, 1997; Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). According to this reductionist view, specific features of a novel string that appeared previously in training strings determines whether the novel string is categorized as grammatical or nongrammatical.

The most common view in the literature on human artificial grammar learning now seems to be that both the holistic, perceptual, rule-learning position *and* the more reductionistic position are needed for an adequate understanding of artificial grammar learning (Dienes, Broadbent, & Berry, 1991; Dulany et al., 1984; Meulemans & Van der Linden, 1997; Perruchet & Pacteau, 1990; Reber, 1989; Servan-Schreiber & Anderson, 1990; Vokey & Brooks, 1992).

We developed a method with which to study artificial grammar learning in nonhuman animals. It may be helpful at the outset to state some disclaimers. First, we offer no comprehensive theory or account of this complex learning, but only the demonstration that it mirrors the analogous human learning in that it involves both simple associative learning and more complex concept learning. Second, one might reasonably ask, Why develop a new task seemingly so complex that no clear, simple, comprehensive description of what is learned can be provided? We believe that *simplicity*, or *parsimony*, despite recent advances in its theoretical meaning (Pitt, Myung, & Zhang, 2002), is still an ambiguous and context-dependent concept with problematic implications for scientific research programs (Chomsky, 1959; Shimp, 1999, 2001) and with a literature (e.g., Nersessian, 1987; Sober, 1975, 1988) not overly familiar to most practicing scientists. In short, we feel it is not yet clear what either *complex* or *simple* means and that, therefore, it might be a mistake to postpone the study of *complex* performances until we have a full understanding of simple performances. Third, the criterion in terms of which we justified our claim that learning involves the learning of concepts was the same two-part operational definition that has been used in the literature to decide that pigeons can learn the concepts of same/different and matching to sample: We asked whether pigeons can achieve above-chance performance on familiar strings *and* on novel transfer strings.

EXPERIMENTAL AND ANALYTICAL METHODS

Animals

Six White Carneaux pigeons (*Columba livia*) were obtained from the Palmetto Pigeon Plant (Sumter, SC). The pigeons were 3–4 years old at the beginning of the experiment and had previously served in experiments on multidimensional category learning involving rectangular stimuli of varying widths and heights (Herbranson et al., 1999). Each was maintained at approximately 80% of its free-feeding weight, with supplemental grain provided as needed in the home cage after daily experimental sessions. The birds were housed individually in standard pigeon cages, with free access to water and grit in a colony room with a 14:10-h light:dark cycle. All experimental sessions took place during the light cycle at approximately the same time, 5–6 days per week.

Apparatus

Five experimental chambers had internal dimensions of approximately 38 × 34.5 × 50 cm (l × w × h). Each chamber had three response keys (3.5 × 3.5 cm) mounted in a horizontal row within a clear Plexiglas window (17 × 7 cm, w × h) in the front wall, 20 cm above the chamber floor. The keys themselves were made of either clear plastic or glass. Stimuli were presented on a 15-in. computer monitor situated approximately 3 cm directly behind the Plexiglas window, in order to facilitate viewing from the chamber (i.e., through the Plexiglas window itself and/or any of the three response keys). The five monitors were interfaced to five personal computers that recorded all data and controlled presentation of stimuli and reinforcement, in the form of access to mixed grain through a hopper located directly below the center key. White noise helped to mask extraneous sounds. (For additional details of the chambers, see Fremouw et al., 1998; Herbranson et al., 1999).

Stimuli

Stimuli were strings of colored letters. The strings varied in length from three to eight letters, which were displayed in text mode on the computer monitor, directly behind the center key, with the string situated horizontally with its center near the center of the screen. The letter set consisted of T, X, V, P, and S, presented in the colors red, white, gray, green, and blue, respectively. These letters were chosen on the basis of their visual discriminability by pigeons (see Blough, 1984). Different letters appeared in different colors in the hope that this additional stimulus dimension would further improve discriminability. We have no reason to believe, however, that the various letter/color pairs, hereafter referred to as *characters*, which varied in geometry, hue, brightness, and probably other dimensions as well, were equally discriminable. This situation mirrors that in artificial grammar learning research with human participants.

Grammatical Strings

Grammatical strings were generated from the finite-state Markovian artificial grammar displayed in Figure 1. Each of the 43 possible grammatical strings, having three to eight characters, was used either as a training stimulus or as a novel transfer stimulus. Table 1 shows the entire set of strings from which strings were presented at various stages of the experiment.

Table 1
Character Strings Used During Training and Transfer

Training Stimuli			
Grammatical		Nongrammatical	
VVS	VXXXXVS	VTS	VXPVXVS
VXVS	VVPXVPS	VVTS	VXXXXPS
TPPTS	VXXXVPS	VTPS	VXXTVPS
TTXVS	TPPPPPTS	TTPPS	VVPPTXVS
VXXVS	TPPTXXVS	TPVVS	TXPTXXVS
TPPPTS	TPTXXXVS	VXVTS	TTXPPXVS
TPTXVS	TTXXXXVS	VPPPTS	TPTXXXXS
TTXXVS	TPPTXVPS	TPXXVS	VPTXXVTS
TTXVPS	TPTXXVPS	TPXXXS	TTXXXVXS
VVPXVS	VXVPXXVS	VVXXXS	VPPXXXVS
VXXXVS	VXXVPXVS	VVXVTS	VXVPXXTS
TPPPPTS	VXXXXXVS	TVPTPTS	TXXVPXTS
TPTXXVS	VVPXXVPS	TVPTXVS	VXXVXXVS
TTXXXVS	VXVPXVPS	TTXVXXS	TXVVXVPS
TPTXVPS	VXXXXVPS	TPTVVVS	VTXVXVPS
VXVPXVS		TPXXVPS	
Transfer Stimuli			
Grammatical		Nongrammatical	
Similar	Nonsimilar	Similar	Nonsimilar
TPTS	TTS	TVS	VPPS
VVPS	VXVPS	TTXVVS	VPTPS
VXXVPS	TPPPTXVS	VVPPVPS	VTVXVS
TPPTXVS	TTXXXVPS	VTVPXVS	VPTXXPS
TTXXVPS	VVPXXXVS	VPPPPPTS	VXPTXVPS
VVPXXVS	TTXVPXVS	TTXXXXTS	TXPXXVPS

Nongrammatical Strings

We generally replicated the way in which the human artificial grammar literature controls for the possibility that various features unrelated to grammatical rules might serve as the basis for discriminating between grammatical and nongrammatical strings. Grammatical and nongrammatical strings were carefully matched in several ways. Nongrammatical strings used the same set of characters as grammatical strings but, as in corresponding human experiments, broke grammatical rules in one or two positions. One nongrammatical string was generated from each possible grammatical string (for a total of 43 nongrammatical strings) by changing either one or two characters from a grammatical string, so that the resulting string could not be legally generated by the grammar. The positions of violations in nongrammatical strings were randomly selected to occur over four positions (beginning with the leftmost character): initial, second, middle, and second to last. (In the case of strings of three letters, violations could occur in the first two locations, and in the case of strings of four letters, violations could occur in all but the last location, and so on.) The violating character was randomly selected from the character set and could appear only in a *legal* position, in the sense that an illegal character never appeared in a position in which it could not be generated by the grammar. Some human experiments using different grammars have included violations in the final position, but this possibility was excluded here since the final character in all possible grammatical strings was S. In addition, S never appeared earlier in a string, since it appeared only in the final position of grammatical strings. Furthermore, only T and V began legal strings, so only those characters appeared in the initial position of strings, regardless of grammatical status. These constraints on the generation of nongrammatical strings were designed to help reduce any tendency for the birds to attend exclusively to single characters in salient locations, since Reber and Lewis (1977) showed that the letters and bigrams in the initial and terminal positions of a string may be particularly salient.

Global String Similarity

Possible bases for grammar learning in humans are string *similarity* and *strength* of component parts of strings (Higham, 1997). Following Vokey and Brooks (1992), similar strings are defined here as those that differ from one another in no more than one position. To investigate the role of overall similarity, half the grammatical transfer strings and half the nongrammatical transfer strings were *similar* (differed in only one position) to at least one string in the training set. The remaining transfer strings were less similar from each training string, in the sense they differed in at least two positions.

Component Parts of Strings

Servan-Schreiber and Anderson (1990) and Meulemans and Van der Linden (1997) suggested that artificial grammar learning could consist in part of learning about permissible components of strings, rather than sets of rules or complete character strings: Grammar learning might not be based exclusively on overall string similarity, as was described above, but on component strength, where the strength of a component is defined as the average number of times each bigram and trigram within it appeared in the training set of grammatical strings. Specifically, the strength of a component bigram or trigram is defined as the average number of times that component appeared in the training set of grammatical strings. For example, the three-character string VVS contains three component bigrams or trigrams; the two bigrams VV and VS and the trigram VVS. Calculating strength for each of these three components involves counting the frequency with which each component appeared in the set of training grammatical strings and averaging across the three (Servan-Schreiber & Anderson, 1990). For this example, VV appeared 4 times, VS 17 times, and VVS 1 time, yielding a strength for the string VVS equal to $(4 + 17 + 1) / 3 = 7.33$.

Strings were selected for transfer so that the difference in strength between similar and nonsimilar transfer strings was small; the strength was 11.7 for similar versus 11.8 for nonsimilar grammatical strings and 8.3 for similar versus 8.6 for nonsimilar nongrammatical transfer strings. Also, since component parts at the beginning or the end of a string may be particularly salient (Knowlton & Squire, 1994; Reber & Lewis, 1977), similar and nonsimilar strings were selected to minimize average strength of bigrams and trigrams in the anchor positions: For initial and final bigrams and trigrams, the strength was 8.6 versus 8.7 for grammatical strings and 5.4 versus 5.5 for nongrammatical strings.

Trial Structure

Trials during training consisted sequentially of an orienting cue, string presentation, a categorization response, either reinforcement or a correction procedure, depending on whether or not the response was correct, and finally, an intertrial interval (ITI). Thirty-one of the grammatical strings and also 31 of the nongrammatical strings appeared during training, with the remaining 12 of each type reserved for use in this transfer test. During training, each of the 31 grammatical strings and each of the 31 nongrammatical strings appeared twice, for a total of 124 trials per day.

Each trial began with an orienting cue (a 2.4 × 2.4 cm green block) presented directly behind the center key. This cue remained on the screen until a bird pecked the center key. Following a peck, the orienting cue was immediately replaced with a character string, as specified above. Half of the strings were grammatical, and the others were nongrammatical. These were ordered randomly each day. For Birds 2, 5, and 6, strings were presented by adding one character at a time to the previously presented characters, beginning from the left. A character was added every 1 sec until all the characters were present. For Birds 3, 4, and 7, all the characters were presented at once, immediately following the center keypeck. We could find no obvious or systematic differences between these two groups of birds. We therefore will treat all 6 birds as a single group. A second peck to the center key following the minimum required observation period of the entire string (2 sec after the addition of the last character for Birds 2, 5, and 6; 5 sec after stimulus presentation for Birds 3, 4, and 7) illuminated a 2.4 × 2.4 cm block behind each side key, red on the left and blue on the right. The bird then categorized the string as grammatical or nongrammatical by pecking one of the side keys. Pecking the red left key was reinforced when the character string was grammatical, and pecking the blue right key was reinforced when the character string was nongrammatical. A character string remained on the screen until a choice response was made. Reinforcement consisted of approximately 2.5-sec access to mixed grain. After reinforcement, a 5-sec ITI preceded the next trial.

If a categorization response was incorrect, a 10-sec correction interval began, during which the houselight flashed on and off every 0.5 sec. This was followed by presentation of the same string, and correction continued until the correct response was given, with no limit to the number of iterations. On a random 10% of daily trials, the consequences of a response, correct or incorrect, were omitted. That is, no reinforcement or correction procedure followed a choice, and the choice was simply followed by an ITI and the next trial. The purpose of this feature was to accustom the birds to the intermittent reinforcement required for subsequent test trials with novel strings, as will be described below.

Criterion for Learning

We could only guess at the beginning of the experiment what level of performance the pigeons might achieve. As was noted above, absolute levels in the corresponding human experiments have varied widely, without any apparent negative implications for the theoretical goal of discriminating between learning concepts (*not learning the grammar*) versus memorizing letter combinations. We therefore selected a criterion for learning equal to 60% correct performance over 5 consecutive days, with the additional constraint that the pattern of performance over those 5 days be nonmonotonic. This criterion was sufficiently high to correspond to a level of performance both reliably greater than chance and adequate to discriminate between learning of abstract concepts and learning of specific letter combinations.

Transfer Test With Novel Strings

Once criterion was reached, we tested the birds' abilities to transfer performance to previously unseen strings. Recall that 12 of both the grammatical and the nongrammatical strings were not used in training and were reserved for use in this transfer test. The grammatical and nongrammatical transfer strings were selected to include two potentially diagnostic characteristics that have been investigated in studies of artificial grammar learning in humans: Specifically, they included strings of (1) all possible string lengths and (2) all possible string types—that is, transfer string lengths varied from three to eight in the same way that training string lengths varied. There was at least one novel grammatical and one novel nongrammatical string of each length. Efforts were made to balance the average lengths of grammatical and nongrammatical transfer strings (6.8 characters for grammatical strings and 6.7 characters for nongrammatical strings). For this particular grammar, there were the following five distinct types of strings that could be generated on the basis of recurring sets of symbols in the grammar (Reber, 1967): (1) T(P)TS, (2) T(P)TX(X)(VPX(X))VS, (3) (P)TX(X)(VPX(X))VPS, (4) V(X)(VPX(X))VS, and (5) V(X)(VPX(X))VPS. Novel transfer strings were balanced in terms of these five string types, with at least two and no more than four of a given string type appearing as either grammatical or nongrammatical strings.

For 5 consecutive days, the 12 novel grammatical strings and the 12 novel nongrammatical strings were randomly included among daily trials. Novel transfer strings were presented in extinction, with no reinforcement or correction, regardless of the response. Instead, following a response, the procedure moved directly to the next trial, after an ITI. Tests were given in extinction to reduce learning of transfer stimuli. Pecks to a random 10% of familiar strings continued not to be followed by either reinforcement or the correction procedure and led directly to the ITI, so that some nonfeedback trials involved novel strings and some involved familiar strings.

Pretraining

Since the birds had already served in a discrete-trials experiment involving visual displays presented on a computer screen, pretraining consisted merely of a few days over which we increased the number of daily trials from about 10 to the experimental value of 124.

RESULTS

Table 2 summarizes categorization performance over the last 10 days of training. Our analysis focuses on the question of whether artificial grammar learning involves learning of concepts, learning of specific letter-string–response associations, or both.

Acquisition

The average number of days of training to criterion was 179, with training for individual birds ranging from 120 to 456 days. Table 2 shows that the average percentage of correct categorizations was 62.3%. This performance is significantly greater than chance performance of 50%: The 95% confidence interval was (59.9, 64.6).

On the average, the birds called more strings grammatical than nongrammatical: The average percentage of categorization responses that were to the left key (the *grammatical* key) was 60.0%, with a resulting 95% confidence interval equal to (50.0, 70.1). It is important to note that this moderate position bias could not have been responsible for the better-than-chance categorization accuracy, since grammatical and nongrammatical trials were equally likely.

Inspection of Table 2 reveals several diagnostic results. For instance, accuracy was high (75.8% correct) on nongrammatical strings of Type 1, relative to accuracy on

Table 2
Percentages of Correct Categorizations During the Final 10 Days of Training

		String Type					String Length						
Bird	Set	1	2	3	4	5	3	4	5	6	7	8	All
2	Nongrammatical	82.5	45.6	60.0	47.5	81.7	80.0	90.0	66.7	54.0	52.5	58.3	59.7
	Grammatical	70.0	71.3	55.0	79.0	84.0	80.0	90.0	73.3	71.7	77.5	70.0	73.5
	All	76.3	57.6	57.5	65.0	82.7	80.0	90.0	70.0	63.6	65.0	64.2	66.6
3	Nongrammatical	77.5	40.0	47.5	26.3	68.3	90.0	85.0	63.3	36.0	43.8	41.7	47.7
	Grammatical	85.0	66.3	85.0	85.0	74.0	90.0	90.0	60.0	75.0	81.3	80.8	78.4
	All	81.3	52.4	66.3	58.9	70.9	90.0	86.7	61.7	57.3	62.5	61.3	63.1
4	Nongrammatical	60.0	45.6	50.0	48.8	75.0	100.0	80.0	80.0	40.0	41.3	55.0	54.5
	Grammatical	72.5	63.8	33.0	72.0	72.0	80.0	70.0	70.0	65.0	63.8	62.5	64.8
	All	66.3	54.1	41.3	61.7	73.6	90.0	76.7	75.0	53.6	52.5	80.0	59.7
5	Nongrammatical	70.0	25.6	42.5	30.0	68.3	100.0	90.0	46.7	32.0	46.3	31.7	42.9
	Grammatical	87.5	81.3	65.0	80.0	84.0	100.0	80.0	80.0	70.0	82.5	81.7	80.0
	All	78.8	51.8	53.8	57.8	75.5	100.0	86.7	63.3	52.7	64.4	56.7	61.5
6	Nongrammatical	70.0	22.2	27.5	25.0	65.0	100.0	90.0	33.3	22.0	33.3	35.0	38.1
	Grammatical	92.5	92.5	90.0	90.0	82.0	80.0	100.0	86.7	81.7	88.8	95.0	90.0
	All	81.3	55.3	58.8	61.1	72.7	90.0	93.3	60.0	54.5	61.3	65.0	63.9
7	Nongrammatical	95.0	62.2	67.5	46.3	93.3	100.0	95.0	83.3	62.0	57.5	69.2	69.0
	Grammatical	60.0	37.5	20.0	58.0	60.0	90.0	50.0	43.3	36.7	47.5	52.5	48.4
	All	77.5	50.6	43.8	52.8	78.2	95.0	80.0	63.3	48.2	52.5	60.8	58.7
Average	Nongrammatical	75.8	40.2	49.2	37.3	75.3	95.0	88.3	62.2	41.0	45.8	48.5	52.0
	Grammatical	77.9	68.8	58.0	77.3	76.0	86.7	80.0	68.9	66.7	73.6	73.8	72.5
	All	76.9	53.6	53.6	59.6	75.6	90.8	85.6	65.6	55.0	59.7	64.7	62.3

nongrammatical strings overall. There are, of course, no articulated theories of avian artificial grammar learning, but intuition suggests that one or two violations in strings of Type 1 (strings such as TPTS, TPPTS, TPPPTS, etc.; see the Method section above) might be expected to be more easily perceived and memorized than violations in more complex string types. Similarly, accuracy was high (95.0% correct) on nongrammatical strings, as well as on grammatical strings (86.7% correct), of length 3. Presumably, both grammatical and nongrammatical strings of such short length are relatively easy to memorize. Perhaps, then, these two kinds of strings, Type 1 and length 3, were categorized accurately because the birds memorized them, rather than because the birds learned anything about concepts or rules.

Novel String Transfer Tests

Table 3 shows percentages of correct categorizations of the 24 novel strings during the 5 days of the transfer test. The birds characterized these novel strings in much the same way as they did the training strings. Overall, 60.7% of the categorizations of novel strings were correct, with the 95% confidence interval equal to (58.3, 63.1).

Table 3 shows categorization accuracy on the two types of strings, Type 1 strings and strings of length 3, that were predicted, on the basis of training performance, to lead to poor performance on novel strings, since original learning seemed to reflect memorization of specific training strings, not learning of general concepts. In fact, as Table 3 shows, categorization accuracy on those two specific kinds of novel strings was not above chance. Specific "easy" strings were apparently memorized during training, so that little or no evidence of concept learning was displayed when the birds were tested with novel strings of those two kinds.

The better-than-chance accuracy to novel test strings means that performance satisfied the second of the two parts of the conventional definition of concept learning. Combined with the results to training strings, this means that both conventional criteria for grammar learning adopted here were satisfied.

Performance to Familiar Strings During the Transfer Test

Overall mean accuracy to familiar strings was 64.3%, with the 95% confidence interval equal to (63.3, 65.3). Mean accuracy to familiar grammatical strings was 71.6%, with the 95% confidence interval equal to (65.6, 77.7). Mean accuracy to familiar nongrammatical strings was 56.9%, with the 95% confidence interval equal to (50.3, 63.5). All these values closely resemble those reported above for original training.

Global Novel String Similarity

Global string similarity involves the number of differing features between a novel transfer string and its most similar counterpart in the training set (Vokey & Brooks, 1992). Similar and nonsimilar strings were balanced with respect to the strengths of their component bigrams and trigrams. Therefore, a direct comparison of categorization of similar and nonsimilar strings was possible without confounding the analysis with the strength of these components. The percentage of correct categorizations of similar and nonsimilar novel strings was 60.6% and 60.8%, respectively. Thus, performance on transfer strings did not depend on whether a string was similar to one or more familiar training strings. It is therefore unlikely that the obtained transfer to novel strings was importantly based on global similarity of transfer strings to training strings.

Table 3
Percentages of Correct Categorizations to Novel Strings During the 5-Day Transfer Test

		String Type					String Length						
Bird	Set	1	2	3	4	5	3	4	5	6	7	8	All
2	Nongrammatical	50.0	70.0	40.0	80.0	50.0	60.0	60.0	40.0	50.0	80.0	60.0	61.7
	Grammatical	60.0	60.0	60.0	60.0	86.7	40.0	80.0	80.0	100.0	60.0	60.0	66.7
	All	55.0	64.0	50.0	73.3	72.0	50.0	73.3	60.0	66.7	70.0	60.0	64.2
3	Nongrammatical	30.0	50.0	40.0	50.0	50.0	20.0	40.0	20.0	60.0	53.3	45.0	45.0
	Grammatical	70.0	60.0	60.0	80.0	80.0	40.0	80.0	80.0	100.0	73.3	60.0	70.0
	All	50.0	56.0	50.0	60.0	68.0	30.0	66.7	50.0	73.3	63.3	52.5	57.5
4	Nongrammatical	30.0	80.0	50.0	40.0	60.0	40.0	20.0	20.0	40.0	53.3	70.0	50.0
	Grammatical	30.0	66.7	80.0	100.0	80.0	0.0	60.0	100.0	80.0	80.0	80.0	71.7
	All	30.0	72.0	65.0	60.0	72.0	20.0	46.7	60.0	53.3	66.7	75.0	60.8
5	Nongrammatical	70.0	60.0	60.0	70.0	20.0	60.0	60.0	80.0	60.0	66.7	45.0	58.3
	Grammatical	20.0	73.3	70.0	60.0	46.7	20.0	30.0	20.0	80.0	80.0	60.0	55.0
	All	45.0	68.0	65.0	66.7	36.0	40.0	40.0	50.0	66.7	48.9	52.5	56.7
6	Nongrammatical	80.0	40.0	30.0	50.0	0.0	40.0	0.0	100.0	50.0	40.0	35.0	41.7
	Grammatical	70.0	80.0	70.0	90.0	93.3	60.0	90.0	80.0	100.0	80.0	80.0	81.7
	All	75.0	64.0	50.0	63.3	56.0	50.0	60.0	90.0	66.7	60.0	57.5	61.7
7	Nongrammatical	50.0	50.0	20.0	65.0	60.0	20.0	80.0	40.0	50.0	66.7	45.0	51.7
	Grammatical	60.0	80.0	70.0	100.0	66.7	60.0	60.0	60.0	80.0	73.3	90.0	75.0
	All	55.0	68.0	45.0	76.7	64.0	40.0	66.7	50.0	60.0	70.0	67.5	63.3
Average	Nongrammatical	51.7	58.3	40.0	59.2	40.0	40.0	43.3	50.0	51.7	60.0	50.0	51.4
	Grammatical	51.7	70.0	68.3	81.7	75.6	36.7	66.7	70.0	90.0	74.4	71.7	70.0
	All	51.7	65.3	54.2	66.7	61.3	38.3	58.9	60.0	64.5	63.2	60.8	60.7

Specific String Types

Accuracy varied in both training and transfer as a function of specific string type. Strings that were most accurately categorized during training (String Types 1 and 5) were also those that were most accurately categorized during transfer, and similarly, strings that were least accurately categorized during training (String Types 2 and 3) were also those that were least accurately categorized during transfer. A two-way (string type and condition) repeated measures analysis of variance (ANOVA) showed a reliable main effect of string type [$F(4,20) = 10.054, p < .001$] and also a reliable main effect of condition type [training, familiar transfer, and novel transfer; $F(2,10) = 10.252, p < .005$].

Accuracy was slightly but reliably better on familiar strings during transfer than during training [$t(4) = 5.81, p < .01$]. It will be recalled that accuracy over the last 10 days of training showed no reliable improvement. Therefore, since accuracy was stable by the end of training, the simple administration of the transfer test, with its novel but unreinforced strings, was apparently sufficient to improve overall accuracy.

Accuracy to novel strings of Type 1 was poorer than that to familiar strings of Type 1 [$t(4) = 5.58, p < .01$], and accuracy to novel Type 2 strings was better than to familiar Type 2 strings [$t(4) = -4.60, p < .01$]. The other three corresponding differences were not reliable. It is possible that strings reserved for transfer of Type 1 were more difficult than training strings, and those of Type 2 were easier. We note, however, that a plausible explanation is to be found in differences in the numbers of exemplars for different string types: During training, the frequencies of different string types were not equal. Specifically, String Types 1, 2, 3, 4, and 5, had 8, 17, 8, 18, and 11 training exemplars, respectively. The correlation between accuracy to novel exemplars of string types and numbers of training exemplars of those types was $r = .95, p < .05$. Thus, poorer transfer to novel Type 1 strings may have been due to training with fewer exemplars of Type 1 strings (Wright, 2001).

DISCUSSION

Multiple Versus Unitary Views of Nonhuman Animal Artificial Grammar Learning

Two main issues in the human artificial grammar learning literature have been, first, whether human participants learn conceptual rules and/or memorize specific letter patterns and, second, whether rule learning is implicit or explicit—that is, whether participants can verbally describe or recognize rules they have learned. Our experiment focused on the first of these two issues, whether pigeons learned *abstract* concepts or remembered specific features. As was reviewed in the introduction, a significant portion of the earlier literature on human artificial grammar learning involved advocates for either side trying to show that the other side was wrong and that their own view was correct. There is a related polarized conflict in the literature on nonhuman animal language and grammar (Herman, 1989; Schusterman & Gisiner, 1989). Smith and Jonides (2000) discussed an analogous situation in the human categorization literature and described two possible positions as the *unitary view* and the *multiple view*. The unitary view is that according to which all categorization is due to just one mechanism. The multiple view acknowledges that categorization may involve multiple mechanisms. Smith and Jonides suggested that from the multiple view, a constructive goal would be to determine the conditions under which different mechanisms apply. Wright (2001) has suggested a similar approach to performance by pigeons in delayed matching-to-sample tasks, where performance is sometimes better described in terms of memories for entire complex stimuli and sometimes better described in terms of memories for individual component features of complex stimuli. Correspondingly, our goal here is not to show that either the holistic or the reductionistic view of artificial grammar learning is correct and the other is wrong but to suggest that both will be required for a complete account of avian artificial grammar learning.

Our data suggest that when shown strings that either do or do not conform to an artificial grammar, pigeons, like humans, can learn to discriminate the one from the other at an above-chance level: (1) Performance was better than chance during training with familiar strings, and (2) performance subsequently remained at approximately the same above-chance level with novel test strings. The combination of these two findings means that performance satisfied the conventional operational definition of concept learning in nonhuman animals. Since the concept under consideration here is an artificial grammar, the results by definition show, in this limited sense, that pigeons can learn an artificial grammar.

We hasten with all due speed, however, to note that this conclusion, which is fully justified on the conventional grounds of an operational definition of concept learning, in our judgment leaves most of the important theoretical questions not only unanswered, but even unaddressed. As was noted in the introduction, showing that performance satisfies this definition does not tell us much, if anything, about the underlying processes by which this performance is achieved.

At the same time that the results revealed this evidence that birds generally learned to do something more than memorize specific strings, the results also showed, with some specific strings, that that is exactly what they did. “Easy” strings of length 3 and strings of Type 1 appeared to be memorized. Memorization of specific strings is understandable, since the birds were given extensive training and since training strings, especially short strings and “easy” strings, were few enough in number for pigeons to memorize (see Vaughan & Greene, 1984; Wright, Cook, Rivera, Sands, & Delius, 1988).

In short, our data correspond better to the *multiple* view of Smith and Jonides (2000) than to the *unitary* view, and this result agrees with the contemporary consensus in the human literature, according to which artificial grammar learning involves both rule learning and associative learning (Knowlton & Squire, 1996).

The Role of Complexity Versus Simplicity in Behavioral Research

If the nature of the learning that takes place in artificial grammar learning is controversial, so too, we suspect, might be the basic nature of the task when used with nonhuman animals. We readily concede the task is *complex*. From a perspective according to which simpler is always better, this task may appear peculiarly and counterproductively complex. Our perspective, in contrast, is that *simple* and *complex* are relative terms and that the present task offers actual advantages of relative simplicity, in comparison, for example, with some naturalistic visual concept learning tasks, in which the dimensions of stimuli belonging to a category—say, *tree*—are unclear in comparison with the clarity with which the rules of an artificial grammar may be specified. In this sense, it is no more complex to ask if a bird learns anything about the rules defining an artificial grammar than it is to ask if a bird learns anything about the abstract visual concept of *tree*. In both cases, it is possible to answer the question affirmatively without knowing exactly what is learned. Thus, we intentionally sacrificed a degree of simplicity to study a more complex problem: How do nonhuman animals learn the logical structure implicit in relatively complex patterns? We hope that the artificial grammar paradigm will prove as useful with nonhuman animals as it has been with human participants.

Complex as artificial grammar learning may be, we feel that it may ultimately prove to be as fruitful for study with nonhuman animals as it has been with humans. We hope, in particular, that future research with this new method will address the problems of what pigeons learn about the rules of an artificial grammar, what conditions induce pigeons to learn abstractions instead of specific features of character strings, and what invasive nonhuman animal neurobiological research will say about the neuropsychological theories that increasingly are the foci of human artificial grammar learning.

REFERENCES

Blough, D. S. (1984). Form recognition in pigeons. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 277-289). Hillsdale, NJ: Erlbaum.

Calvin, W. H., & Bickerton, D. (2000). *Lingua ex machina: Reconciling Darwin and Chomsky with the human brain.* Cambridge, MA: MIT Press.

Candland, D. K. (1993). *Feral children and clever animals: Reflections on human nature.* Oxford: Oxford University Press.

Catania, A. C. (1972). Chomsky's formal analysis of natural languages: A behavioral translation. *Behaviorism*, **1**, 1-15.

Catania, A. C. (1991). The phylogeny and ontogeny of language function. In N. A. Krasnegor & D. M. Rumbaugh (Eds.), *Biological and behavioral determinants of language development* (pp. 263-285). Hillsdale, NJ: Erlbaum.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1959). Review of B. F. Skinner: *Verbal behavior. Language*, **35**, 26-58.

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

Cook, R. G., Cavoto, K. K., & Cavoto, B. R. (1995). Same–different texture discrimination and concept learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **21**, 253-260.

Cook, R. G., Katz, J. S., & Cavoto, B. R. (1997). Pigeon same–different concept learning with multiple stimulus classes. *Journal of Experimental Psychology: Animal Behavior Processes*, **23**, 417-433.

Crowther-Keyck, H. (1999). George A. Miller, language, and the computer metaphor of mind. *History of Psychology*, **2**, 37-64.

Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **5**, 1322-1338.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 875-887.

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.

Fremouw, T., Herbranson, W. T., & Shimp, C. P. (1998). Priming of attention to local or global levels of visual analysis. *Journal of Experimental Psychology: Animal Behavior Processes*, **24**, 278-290.

Fremouw, T., Herbranson, W. T., & Shimp, C. P. (2002). Dynamic shifts of avian local/global attention. *Animal Cognition*, **5**, 233-243.

Gardner, R. A., Gardner, B. T., & Van Cantfort, T. E. (Eds.) (1989). *Teaching sign language to chimpanzees.* Albany: State University of New York Press.

Herbranson, W. T. (2000). *Cognitive dissociation of memory systems: Implicit and explicit memory in pigeons* (Columba livia). Unpublished doctoral dissertation, University of Utah.

Herbranson, W. T., Fremouw, T., & Shimp, C. P. (1999). The randomization procedure in the study of categorization of multi-dimensional stimuli by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **25**, 113-135.

Herbranson, W. T., Fremouw, T., & Shimp, C. P. (2002). Categorizing a moving target in terms of its speed, direction, or both. *Journal of the Experimental Analysis of Behavior*, **78**, 249-270.

Herman, L. M. (1989). In which procrustean bed does the sea lion sleep tonight? *Psychological Record*, **39**, 19-50.

Herman, L. M., & Uyeyama, R. K. (1999). The dolphin's grammatical competency: Comments on Kako (1999). *Animal Learning & Behavior*, **27**, 18-23.

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science*, **146**, 549-551.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **2**, 285-301.

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

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar.* Cambridge, MA: MIT Press.

Kako, E. (1999). Elements of syntax in the systems of three language-trained animals. *Animal Learning & Behavior*, **27**, 1-14.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & 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, & Cognition*, **22**, 169-181.

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

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

Nersessian, N. J. (Ed.) (1987). *The process of science: Contemporary philosophical approaches to understanding scientific practice.* Dordrecht: Nijhoff.

Pepperberg, I. M. (1999a). *The Alex studies: Cognitive and communicative abilities of grey parrots.* Cambridge, MA: Harvard University Press.

Pepperberg, I. M. (1999b). Rethinking syntax: A commentary on E. Kako's

"Elements of syntax in the systems of three language-trained animals." *Animal Learning & Behavior*, **27**, 15-17.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, **119**, 264-275.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, **109**, 472-491.

Premack, D. (1986). *Gavagai! Or the future history of the animal language controversy*. Cambridge, MA: MIT Press, Bradford Books.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & 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., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition*, **5**, 333-361.

Roitblat, H. L., Harley, H. E., & Helweg, D. A. (1993). Cognitive processing in artificial language research. In H. L. Roitblat, L. M. Herman, & P. E. Nachtigall (Eds.), *Language and communication: Comparative perspectives* (pp. 1-23). Hillsdale, NJ: Erlbaum.

Savage-Rumbaugh, S., Shanker, S. G., & Taylor, T. J. (1998). *Apes, language, and the human mind*. New York: Oxford University Press.

Schusterman, R. J., & Gisiner, R. C. (1989). Please parse the sentence: Animal cognition in the procrustean bed of linguistics. *Psychological Record*, **39**, 3-18.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 592-608.

Shimp, C. P. (1999). Tolerance in a rigorous science. *Journal of the Experimental Analysis of Behavior*, **71**, 284-288.

Shimp, C. P. (2001). Behavior as a social construction. *Behavioural Processes*, **54**, 11-32.

Shimp, C. P., Herbranson, W. T., & Fremouw, T. (2001). Avian visual attention in science and culture. In R. G. Cook (Ed.), *Avian visual cognition* [On line.] Available at www.pigeon.psy.tufts.edu/avc/shimp.

Smith, E. E., & Jonides, J. (2000). The cognitive neuroscience of categorization. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.). Cambridge, MA: MIT Press.

Sober, E. (1975). *Simplicity*. Oxford: Oxford University Press, Clarendon Press.

Sober, E. (1988). *Reconstructing the past: Parsimony, evolution, and inference*. Cambridge, MA: MIT Press.

Terrace, H. S. (1985). In the beginning was the "Name." *American Psychologist*, **40**, 1011-1028.

Vaughan, W., Jr., & Greene, L. L. (1984). Pigeon visual memory capacity. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 256-271.

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.

Wasserman, E. A., Hugart, J. A., & Kirkpatrick-Steger, K. (1995). Pigeons show same–different conceptualization after training with complex visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, **21**, 248-252.

Wright, A. A. (1997). Concept learning and learning strategies. *Psychological Science*, **8**, 119-123.

Wright, A. A. (2001). Learning strategies in matching to sample. In R. G. Cook (Ed.), *Avian visual cognition* [On line.] Available at www.pigeon.psy.tufts.edu/avc/wright.

Wright, A. A., Cook, R. G., Rivera, J. J., Sands, S. F., & Delius, J. D. (1988). Concept learning by pigeons: Matching-to-sample with trial-unique video picture stimuli. *Animal Learning & Behavior*, **16**, 436-444.

NOTE

1. Since the structure in artificial grammar tasks is purely formal, we feel that it better resembles orthographic regularity, or the statistical structure of natural language, than it does naturalistic grammar. We are therefore inclined to believe that it is arbitrary whether these tasks are called *artificial grammar* tasks or *artificial orthographic regularity* tasks. In this sense, the present research is quite different from research on language or communication in dolphins, parrots, chimpanzees, and other nonhuman animals, as well as in humans (Calvin & Bickerton, 2000; Candland, 1993; Catania, 1972, 1991; Gardner, Gardner, & Van Cantfort, 1989; Herman, 1989; Herman & Uyeyama, 1999; Hirsh-Pasek & Golinkoff, 1996; Kako, 1999; Pepperberg, 1999a, 1999b; Premack, 1986; Savage-Rumbaugh, Shanker, & Taylor, 1998; Terrace, 1985).

In short, we make no claims about the relation between artificial and naturalistic grammar. In the present paper, words like *grammar* or *grammatical* should be understood to be shorthand for *artificial grammar* or *grammatical with respect to an artificial grammar*. This terminology is less than ideal, because it might cause an unwary reader to believe that the present paper necessarily has something to do with naturalistic grammar (as pointed out in an analogous case by, e.g., Schusterman & Gisiner, 1989). We feel that this potential problem is outweighed, however, by an even greater potential problem that might arise if a new term were coined for the present task. Any new term might (1) unduly minimize the potential relations between the present research on nonhuman animal cognition and the human research on which it is based, (2) arbitrarily distance the present empirical research from nonhuman animal conceptual literature that has already used the traditional term *synthetic grammar* (Roitblat, Harley, & Helweg, 1993), and (3) merely introduce a new, arbitrary term for the well-established *artificial* or *synthetic* grammar.

(Manuscript received April 24, 2002;
revision accepted for publication November 3, 2002.)