# Artificial grammar learning in pigeons

**WALTER T. HERBRANSON**
*Whitman College, Walla Walla, Washington*

AND

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

In two experiments, we developed a new methodology for studying complex stimulus control by spatial sequences of letters generated by artificial grammars. An artificial grammar is a system of rules that defines which letter sequences or strings are "grammatical." In Experiment 1, pigeons learned to respond differently to strings conforming to a grammar versus strings that were nongrammatical distortions. Several different criteria all suggested that performance was controlled both by some short chunks of strings shared between reinforced training strings and novel transfer strings and by more complex sequential regularities. In Experiment 2, pigeons quickly and accurately learned to respond differently to strings conforming to one or the other of two different artificial grammars. As in Experiment 1, performance was controlled both by some short chunks and by more complex sequential regularities. The results are interpreted in terms of *family resemblance* and pose new goals for theories of complex stimulus control.

Categorization performances by pigeons can be controlled by such complex spatial regularities across exemplars that explanations based on simple logical relations among stimuli become problematic (Herrnstein, Loveland, & Cable, 1976). Pigeons learn to discriminate between images of fish and nonfish, between images of human figures and nonhuman figures, and so on, without our knowing what regularities across exemplars of different fish or different human figures form the basis for the discriminations. Wittgenstein (1953) applied the term *family resemblance* to describe this kind of naturalistic discrimination in humans, and the term has subsequently been used to describe experimental results from both humans and pigeons (Herrnstein et al., 1976; Rosch, 1978). Wittgenstein used *family resemblance* not just to describe how humans perceive visual similarities across different members of the same family, however, but also to relate human naturalistic visual perception and human natural language grammar. In both cases, there is no known set of defining regularities across exemplars, yet humans learn the distinction between grammatical and nongrammatical, as well as the distinction between the visual appearances of the Smiths and the Joneses.

Chomsky and Miller (1958) developed artificial grammars to facilitate understanding the regularities that permit humans to discriminate between grammatical and nongrammatical. An artificial grammar defines the rules that assign exemplars to "grammatical" categories. Chomsky and Miller may have briefly believed that artificial grammars could be viewed as simple and direct models

of natural language grammars, but few if any researchers now believe that the relation between learning artificial and natural language grammars is either simple or direct (Gomez & Gerken, 1999; Meulemans & Van der Linden, 1997; Servan-Schreiber & Anderson, 1990). Some researchers advocate a position according to which humans learn artificial grammars in a manner that involves component processes the understanding of which might facilitate understanding the evolution of natural language (Gentner, Fenn, Margoliash, & Nusbaum, 2006; Hauser, Chomsky, & Fitch, 2002).

We were therefore curious to know whether any nonhuman species could learn an artificial grammar, and we accordingly developed a pigeon version (Herbranson, 2000; Herbranson & Shimp, 2003) of the artificial grammar learning (AGL) task invented by Reber (1967) on the basis of the ideas introduced by Chomsky and Miller (1958). An artificial grammar is a system of rules that defines "grammatical" strings of letters. One such grammar is displayed in the top panel of Figure 1. This grammar is such that one state depends only on the immediately preceding state. Grammatical strings are generated by starting at the leftmost arrow and by moving through the grammar according to the rules, indicated by the arrows, with each transition adding a letter to the string, until the process finishes at the rightmost arrow. In our initial AGL experiment (Herbranson & Shimp, 2003), pigeons viewed strings of letters that conformed to the grammar shown at the top of Figure 1 or that violated it at one or two locations so that they were nongrammatical distortions. After

---

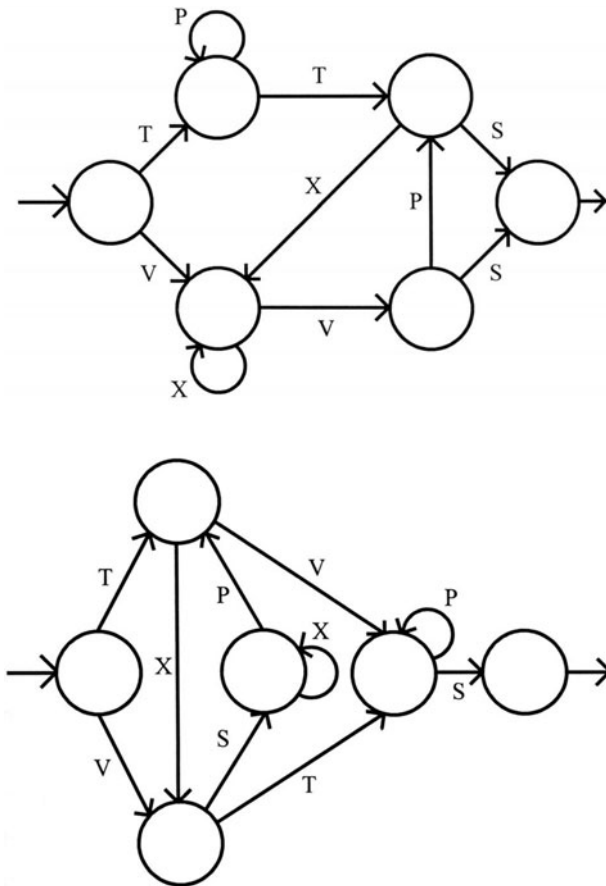**C. P. Shimp, charlie.shimp@psych.utah.edu**

**Figure 1. Top: The artificial grammar used by Reber (1967) and by Herbranson and Shimp (2003) and as one of the two artificial grammars in Experiment 2 (specifically, Grammar 1). Bottom: The artificial grammar used in Experiment 1. From "Finite-State Languages," by N. Chomsky and G. A. Miller, 1958, *Information and Control*, *1*, p. 102, copyright 1958 by Elsevier, Inc., and "Transfer of Syntactic Structure in Synthetic Languages," by A. S. Reber, 1969, *Journal of Experimental Psychology*, *81*, p. 116, copyright 1969 by the American Psychological Association. Adapted with permission.**

viewing a string, a pigeon was required to peck a left key if the string conformed to the grammar and a right key if it did not. After extended training and to a not very high level of accuracy, average performance of 6 pigeons was above chance for the specific strings viewed during training and also, during a transfer test, for novel strings never previously viewed. These results were compatible with pigeons' having a capability to discriminate between regularities across exemplars produced by an artificial grammar and systematic distortions of those regularities. From the perspective of family resemblance, pigeons perceived grammatical strings as belonging to a *family* and nongrammatical strings as belonging to a different family or to none at all.

Converging evidence compatible with a capacity for AGL in birds was recently obtained by Gentner et al. (2006), who taught European starlings to recognize songs composed of two of their basic perceptual units, *rattles* and *warbles*. Gentner et al. sequentially arranged these two mo-

tifs, call them A and B, according to a simple recursive rule generating strings AB, AABB, AAABBB, and so on, with each successive AB pair inserted at the middle of a previous string. They found that after training, most of their starlings could recognize novel songs conforming to this rule, as opposed to the rule ABAB, ABABAB, and so on. The results of Gentner et al. therefore were consistent with the possibility that an avian species can learn to recognize regularities across sequential stimuli generated by a rule.

## EXPERIMENT 1

Herbranson and Shimp (2003) used an actual artificial grammar more complex than a single recursive rule and found evidence compatible with the possibility that pigeons learned complex sequential regularities, as well as short fragments, reinforced during training and present in some novel strings. If it were generally true that pigeons have a capacity for AGL, a new comparative tool would become available with which to study avian complex categorization performances. AGL would provide a way to study control by sequential regularities that are highly complex, yet not so complex as to be unknown, as they are in natural language grammar. Indeed, through the use of a suitable artificial grammar, an experimenter would have full control over the complex regularities. In Experiment 1, we therefore further developed empirical and analytical methods for studying AGL in pigeons.

### Method

**Animals**. Four white Carneaux pigeons (*Columba livia*) were obtained from the Palmetto Pigeon Plant (Sumter, SC). Each was maintained at approximately 80% of free-feeding weight, with supplemental grain provided as needed in their 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 the experimental sessions took place during the light cycle at approximately the same time each day, 5–6 days per week. All 4 birds were experimentally naive.

**Apparatus**. The four experimental chambers had internal dimensions of 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 floor. The keys themselves were made of either clear plastic or glass. The stimuli were presented on a 14-in. computer monitor situated 3 cm directly behind the Plexiglas window. Each monitor was interfaced to its own personal computer 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.

**Stimuli**. The stimuli were horizontal strings of colored letters. Strings varied in length from three to eight letters. The capital letters T, X, V, P, and S were displayed in text mode on the computer monitor, directly behind the center key, in the colors red, white, gray, green, and blue, respectively. These letters were chosen on the basis of their visual discriminability for pigeons (Blough, 1984).

**Grammatical strings**. Grammatical strings were generated from the Markovian grammar displayed at the bottom of Figure 1. This grammar originally appeared in an early demonstration of human AGL (Reber, 1969) and has since been widely used in research on AGL. Strings are formed by starting at the left arrow and moving through the diagram from state to state along the arrows. Each transition between states adds a letter to the string, until the exit via the right arrow. Each of the 43 possible grammatical strings, having between three and eight letters, was used either as a training stimulus

or as a transfer stimulus. Appendix A shows all the strings that were used in Experiment 1.

**Nongrammatical strings**. Nongrammatical strings were distortions of grammatical strings. They used the same letters as grammatical strings but broke the rules of the grammar in either 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 letters from the string to another from the set, so that the resulting string could not be legally generated by the grammar. The position of each violation in a nongrammatical string was randomly selected to occur in one of four positions (beginning with the leftmost letter): initial, second, middle, or second to last. The violating letter was randomly selected from the vocabulary, with the exception that a letter could appear only in a "legal" position (i.e., a letter did not appear in a position in which it could never be generated by the grammar). Since only the letters T and V began legal strings, only those letters appeared in the initial position of stimuli, regardless of grammatical status. Reber and Lewis (1977) showed that letters in the initial and terminal positions of a letter string may be particularly salient, suggesting that these constraints may be especially important. The various limitations described above were, therefore, imposed on the generation of nongrammatical letter strings to reduce the likelihood that a bird would use response strategies based on single letters that might interfere with learning of more complex regularities.

**Novel grammatical and nongrammatical strings**. Thirty-one of the grammatical and 31 of the nongrammatical strings were used during training, with the remaining 12 of each type retained for use in transfer tests. The grammatical and nongrammatical transfer stimuli were selected to include a broad range of potentially diagnostic characteristics that have been investigated in studies of grammar learning in humans. Specifically, transfer stimuli included strings containing all possible string lengths and string types. Like the training stimuli, string lengths of the transfer stimuli varied from 3 to 8 letters, with at least 1 grammatical and 1 nongrammatical string of each length reserved for the transfer test. Furthermore, the average lengths of grammatical and nongrammatical transfer stimuli were approximately equal (6.25 letters for grammatical strings and 6.17 for nongrammatical strings).

**Trial structure**. A daily training session consisted of 62 trials. Each trial during training consisted sequentially of an orienting cue, a letter string, and a response, followed by either reinforcement or a correction procedure, depending on whether or not the response was correct, and, finally, an intertrial interval.

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 center keypeck, the orienting cue was immediately replaced with a letter string, as specified above. Half of the letter strings were grammatical, and the other half were nongrammatical, presented in random order. The first peck to the center key following a minimum 5-sec observation period illuminated a 2.4 × 2.4 cm block behind each side key. A bird then pecked a side key. Pecking the red left key was reinforced when the letter string was grammatical, and pecking the blue right key was reinforced when the letter string was nongrammatical. A letter string remained on the screen until a choice response was made. Reinforcement consisted of approximately 2.5-sec access to mixed grain through the chamber's food hopper (this time varied slightly among birds in order to maintain individual deprivation levels). After reinforcement, a 5-sec intertrial interval preceded the next trial.

An incorrect side key response initiated a 10-sec correction interval 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, either correct or incorrect, were omitted. That is, a choice produced neither reinforcement nor a correction procedure, regardless of its accuracy, and the choice was simply followed by an intertrial interval and the next trial. This procedure

was designed to accustom the birds to trials without reinforcers for correct responses because, in the subsequent transfer test, all trials with novel stimuli went unreinforced (as will be described below).

**Criterion for learning**. The criterion for learning was based on performance in Herbranson and Shimp (2003) and was set at a minimum 60% correct over each of 5 consecutive days, with the additional constraint that the day-to-day pattern of performance over those 5 days be nonmonotonic (e.g., accuracy could not successively get either better or worse on each day).

**Transfer test with novel strings**. Once the learning criterion was reached, the birds' ability to transfer training to previously unseen strings was tested. For 10 consecutive sessions, the remaining 12 unused grammatical strings and a set of 12 previously unseen nongrammatical strings were randomly included among the daily trials that included strings from the 62 training strings. As is usual in generalization tests, 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, following an intertrial interval.

**Definition of the strength of a string fragment, or chunk**. AGL might consist, in part or entirely, of learning small fragments, or chunks of strings, rather than of learning more complex sequential regularities across longer parts of strings (Meulemans & Van der Linden, 1997; Servan-Schreiber & Anderson, 1990). We developed a definition of the *strength* of a small chunk to discriminate between these two possibilities. The definition reflected how a chunk appeared in both grammatical and nongrammatical training strings. Specifically, the strength of a chunk was defined as the number of occurrences of that chunk in grammatical training strings minus the number of its occurrences in nongrammatical training strings. Thus, a chunk with a positive strength was more likely to occur in a grammatical training string, and one with a negative strength was more likely to occur in a nongrammatical training string. The absolute value of a chunk's strength reflects the diagnosticity of that chunk, with a chunk having a strength of zero being completely ambiguous, having occurred equally often in grammatical and nongrammatical training strings.

An example clarifies how chunk strength was calculated. Consider the string VTS. This string contained three chunks: the two bigrams VT and TS and the trigram VTS. The calculation of strength for the string VTS involved counting the frequency with which each bigram and trigram appeared in grammatical and nongrammatical training strings, subtracting the latter from the former, and averaging across all three components. For this example, VT appeared five times in grammatical strings and three times in nongrammatical strings, for a difference of 2. The bigram TS appeared five times in grammatical strings and eight times in nongrammatical strings for a difference of $-3$. The trigram VTS appeared once in a grammatical string and once in nongrammatical strings, for a difference of 0. Averaging these three numbers yielded a strength for VTS of $(2 - 3 + 0)/3 = -0.33$.

## Results

**Acquisition**. Three of the 4 birds reached criterion in 114, 152, and 187 days of training, respectively, including pretraining. One bird did not reach criterion in 269 days and was discontinued from the experiment. Performance over the last 10 days of training averaged 66% ($SD = 4.7$) correct across the 3 birds. This performance was significantly greater than chance performance of 50%, with a 95% confidence interval (CI) of (61%–69%); effect size $r$ (Rosnow & Rosenthal, 2003, p. 231) was equal to a high .971. Average performance was not dissimilar to the average of 62% in Herbranson and Shimp (2003) with different birds and a different grammar. This level of performance was relatively low but was within the range of performances in some human AGL experiments. For example, Reber (1967) obtained an average correct performance of
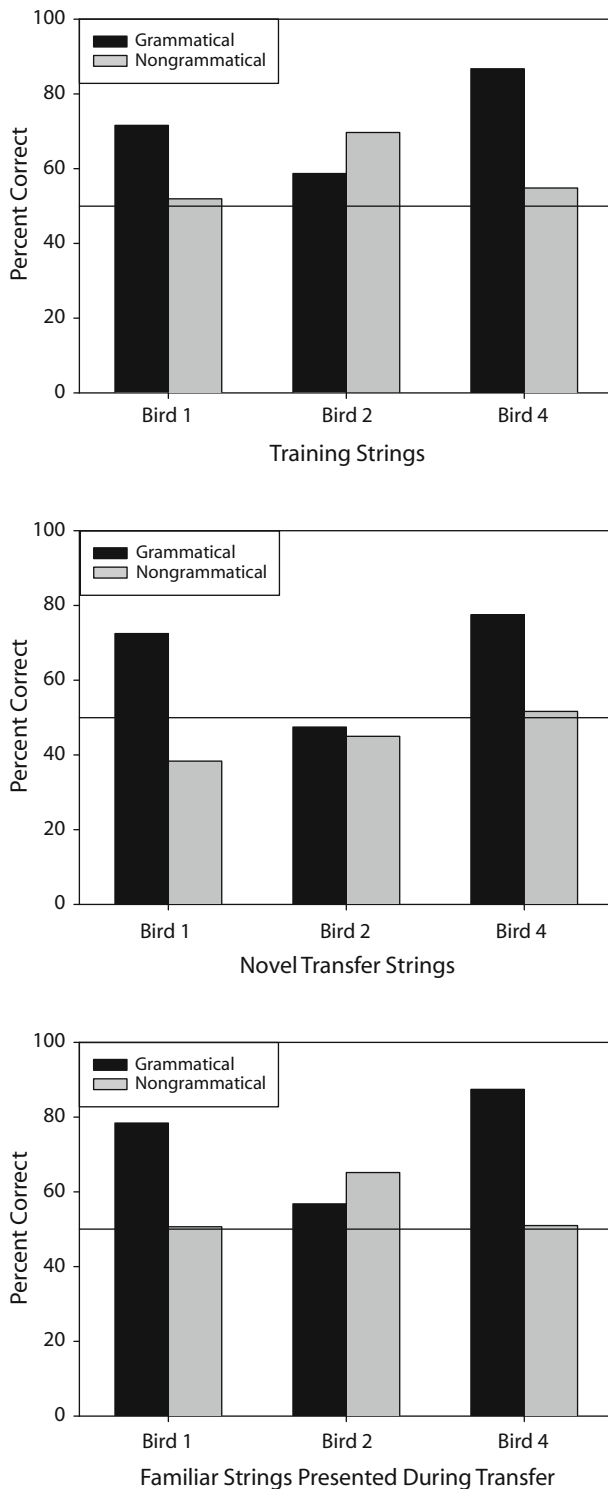
**Figure 2. Top: Percentages of correct responses to grammatical (dark bars) and nongrammatical (light bars) strings over the last 10 days of training for all 3 birds in Experiment 1. Middle: Percentages of correct responses to novel grammatical (dark bars) and novel nongrammatical (light bars) strings over 10 days of transfer for all 3 birds in Experiment 1. Bottom: Percentages of correct responses to familiar grammatical (dark bars) and familiar nongrammatical (light bars) strings over 10 days of transfer for all 3 birds in Experiment 1. In all three panels, the horizontal reference line indicates chance performance of 50%.**

79%, and McAndrews and Moscovitch (1985) obtained 57% correct performance, which, while low, was still reliably greater than chance. More telling than the absolute level of performance was that the effect size was large.

Overall, accuracy was fairly similar for all 3 birds (62%, 64%, and 70% for Birds 1, 2, and 4, respectively). These similar averages concealed, however, some individual differences with important theoretical implications. The top panel of Figure 2 shows that all 3 birds tended to respond accurately to grammatical string but that only Bird 2 responded accurately to nongrammatical strings.

Birds 1 and 4 correctly responded to 80% ($SD = 10.7$) of the grammatical strings but only 53% ($SD = 2.1$) of the nongrammatical strings. Thus, for these 2 birds, performance on grammatical strings was notably higher (95% CI, 69%–90%; effect size $r$ equal to a large .968) than performance on nongrammatical strings (95% CI, 51%–55%; effect size $r$ equal to a large .919). This difference defined a bias for the key corresponding to grammatical strings: Birds 1 and 4 chose the left, grammatical key more often than the right, nongrammatical key. Specifically, they made 64% ($SD = 4.3$) left pecks. This preference by Birds 1 and 4 for the grammatical key replicated the corresponding bias by all 6 birds in Herbranson and Shimp (2003). Across the two experiments, 8 of 9 birds preferred the key associated with grammatical strings. (This preference cannot explain the reliable AGL obtained here, because grammatical and nongrammatical strings were equally likely. For example, if a bird learned nothing about the two categories and responded on the left a random 64% of the time, it would have been correct to 64% of the grammatical strings and incorrect to 64% of the nongrammatical strings, whereas the birds were actually correct to 80% of the grammatical strings and were incorrect to only 47% of the nongrammatical strings. To repeat, the position bias cannot account for the data.)

Bird 2's performance was different in at least two ways. Its accuracy on nongrammatical strings was the highest of the 3 birds, and in fact, its accuracy was actually higher on nongrammatical strings than on grammatical strings (59% and 70% accuracy for grammatical and nongrammatical strings, respectively). Bird 2 also took the longest of the 3 birds to learn the task and reached criterion only after 187 days of training, as compared with 152 and 114 days for Birds 1 and 4. Therefore, Bird 2 may have learned the task in a qualitatively different way, and its data are correspondingly presented separately for all the subsequent analyses.

The results summarized in the top panel of Figure 2 can be interpreted in terms of family resemblance: Birds 1 and 4 learned to perceive regularities across grammatical strings as defining a category, or family, and their having learned the category facilitated their more accurately categorizing grammatical strings. Inaccurate performance by these birds to nongrammatical strings may be interpreted as their not having learned to perceive the regularities and distortions in nongrammatical strings as a category. Bird 2, however, may not have learned a category at all.

Different types of strings might be learned differently (Reber, 1967). Strings can be typed according to the paths through the grammar that produced them. Four such types

**Table 1**
**Percentages of Correct Responses During Training in Experiment 1**

| | String Type | | | | String Length | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | 1 | 2 | 3 | 4 | 3 | 4 | 5 | 6 | 7 | 8 |
| **Bird 1** | | | | | | | | | | |
| Grammatical | 72.0 | 86.0 | 60.0 | 75.7 | 100.0 | 85.0 | 75.0 | 62.0 | 72.2 | 71.7 |
| Nongrammatical | 38.8 | 51.7 | 51.3 | 64.4 | 30.0 | 40.0 | 70.0 | 68.0 | 63.8 | 39.2 |
| Overall | 57.2 | 67.3 | 55.9 | 69.4 | 55.0 | 62.5 | 72.5 | 65.0 | 68.2 | 54.8 |
| **Bird 2** | | | | | | | | | | |
| Grammatical | 61.0 | 68.0 | 54.4 | 54.3 | 40.0 | 30.0 | 80.0 | 64.0 | 57.8 | 60.0 |
| Nongrammatical | 72.5 | 73.3 | 70.0 | 64.4 | 80.0 | 75.0 | 95.0 | 82.0 | 81.3 | 52.3 |
| Overall | 66.1 | 70.9 | 61.8 | 60.0 | 60.0 | 52.5 | 87.5 | 73.0 | 68.8 | 56.0 |
| **Bird 4** | | | | | | | | | | |
| Grammatical | 93.0 | 98.0 | 82.2 | 75.7 | 100.0 | 100.0 | 100.0 | 84.0 | 87.8 | 81.7 |
| Nongrammatical | 51.3 | 56.7 | 61.3 | 51.1 | 50.0 | 45.0 | 65.0 | 70.0 | 63.8 | 43.9 |
| Overall | 74.4 | 75.5 | 72.4 | 61.9 | 75.0 | 72.5 | 82.5 | 77.0 | 76.5 | 62.0 |
| **Average** | | | | | | | | | | |
| Grammatical | 75.3 | 84.0 | 65.6 | 68.6 | 80.0 | 71.7 | 85.0 | 70.0 | 72.6 | 71.1 |
| Nongrammatical | 54.2 | 60.6 | 60.8 | 60.0 | 53.3 | 53.3 | 76.7 | 73.3 | 69.6 | 45.1 |
| Overall | 65.9 | 71.2 | 63.3 | 63.8 | 63.3 | 62.5 | 80.8 | 71.7 | 71.2 | 57.6 |

can be constructed from the present grammar (parentheses indicate recursive loops that can be repeated):

1. T(XS(X)P)V(P)S

2. TXT(P)S

3. V(S(X)PX)T(P)S

4. VS(X)PV(P)S

Birds might learn some types of strings but not others, on the basis of the presence or absence of specific orthographic patterns that result from recurring loops in the grammar. If so, birds might not learn complex regularities but might learn to search for short visual features or chunks. We therefore determined whether performance was better for some types of strings than for others.

Table 1 shows percentages of correct responses as a function of grammatical versus nongrammatical string and of string type and string length, averaged over the last 10 days of training. For all 3 birds, performance on every grammatical string type was numerically above 50%, although several of these numbers are based on frequencies too small to permit tests for statistical reliability. Table 1 shows that grammatical strings of String Type 2 produced the highest levels of performance of any string type for each of the 3 birds. String Type 2 seems intuitively to be the "simplest" of the four types because it contained only a single, one-letter loop. The average lengths of training stimuli of the four types were as follows: Type 1, 6.67; Type 2, 6.55; Type 3, 6.65; and Type 4, 7.25. Thus, string length varied little as a function of string type: Even Type 4 strings, on the average the longest, were only about 10% longer than Type 2 strings, on the average the shortest. Any effect of string type on responding was, therefore, not likely to have been attributable to average string length.

Table 1 also shows another way in which different strings were processed differently. Birds 1 and 4 were perfect on the shortest grammatical strings. Intuitively, the shortest strings would seem easiest to learn on a specific string-by-string basis.

Table 1 suggests still other ways in which different strings may have been processed differently, but these two examples, grammatical strings of Type 2 and of the shortest length, suffice to suggest that at least some short chunks and simple strings might have been memorized.

**Novel string transfer**. The traditional criterion for establishing that nonhuman animals can learn complex regularities is the demonstration that training generalizes to novel stimuli. We therefore asked how the birds responded to previously unseen strings presented in extinction after training. The middle panel of Figure 2 displays the average results for all 3 birds for transfer to the 24 novel strings. Birds 1 and 4 responded to novel strings in much the same way they did training strings. Overall, 60% ($SD = 6.5$) of novel strings were responded to correctly as either grammatical or nongrammatical (95% CI, 56%–65%; effect size $r$ equal to a large .909). Thus, performance was reliably better than chance, and despite the fact that performance was not high on an absolute basis, the effect size was large.

The response bias seen during training continued into novel transfer trials; 65% ($SD = 2.9$) of responses by Birds 1 and 4 were on the left key, so that accuracy continued to be higher on grammatical strings than on nongrammatical strings. Specifically, 75% ($SD = 3.5$) of grammatical strings were correctly classified (95% CI, 73%–77%; effect size $r = .995$), whereas only 45% ($SD = 9.4$) of nongrammatical strings were correctly classified, which was indistinguishable from chance (95% CI, 38%–52%; effect size $r = .599$). Again, these birds learned to respond more accurately to grammatical strings than to nongrammatical strings.

It will be recalled that Bird 2's performance was different during training. It remained so during transfer to novel strings: Unlike Birds 1 and 4, Bird 2 failed to transfer training performance to novel test strings. Bird 2 correctly categorized only 45% of grammatical strings and only 48% of nongrammatical strings, yielding an overall 46% accuracy for transfer, with chance being, of course, 50%. Thus, by the conventional operational definition, Bird 2 failed to learn any complex regularities. The simplest explanation appears to be that Bird 2 memorized simple short strings or chunks

**Table 2**
**Percentages of Correct Responses During Transfer in Experiment 1**

|  | String Type | | | | String Length | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
|  | 1 | 2 | 3 | 4 | 3 | 4 | 5 | 6 | 7 | 8 |
| Bird 1 | | | | | | | | | | |
| Grammatical | 100.0 | 76.7 | 73.3 | 55.0 | 100.0 | 100.0 | 60.0 | 55.0 | 70.0 | 75.0 |
| Nongrammatical | 45.0 | 35.0 | 37.5 | 30.0 | 20.0 | 0.0 | 35.0 | 65.0 | 53.3 | 26.7 |
| Overall | 63.3 | 60.0 | 52.9 | 46.7 | 60.0 | 50.0 | 47.5 | 60.0 | 60.0 | 54.3 |
| Bird 2 | | | | | | | | | | |
| Grammatical | 25.0 | 30.0 | 63.0 | 60.0 | 50.0 | 60.0 | 60.0 | 25.0 | 33.3 | 53.3 |
| Nongrammatical | 47.5 | 40.0 | 42.5 | 60.0 | 20.0 | 30.0 | 60.0 | 60.0 | 50.0 | 45.0 |
| Overall | 40.0 | 34.0 | 51.4 | 60.0 | 35.0 | 45.0 | 60.0 | 42.5 | 40.0 | 48.6 |
| Bird 4 | | | | | | | | | | |
| Grammatical | 45.0 | 70.0 | 90.0 | 90.0 | 30.0 | 60.0 | 100.0 | 80.0 | 85.0 | 77.5 |
| Nongrammatical | 40.0 | 50.0 | 67.5 | 45.0 | 70.0 | 60.0 | 45.0 | 75.0 | 43.3 | 40.0 |
| Overall | 41.7 | 62.0 | 77.1 | 75.0 | 50.0 | 60.0 | 72.5 | 77.5 | 60.0 | 61.4 |
| Average | | | | | | | | | | |
| Grammatical | 56.7 | 58.9 | 75.4 | 68.3 | 60.0 | 73.3 | 73.3 | 53.3 | 62.8 | 68.6 |
| Nongrammatical | 44.2 | 41.7 | 49.2 | 45.0 | 36.7 | 30.0 | 46.7 | 66.7 | 48.9 | 37.2 |
| Overall | 48.3 | 52.0 | 60.5 | 60.6 | 48.3 | 51.7 | 60.0 | 60.0 | 53.3 | 54.8 |

of strings during training and learned little or nothing else, in a manner consistent with our previous interpretation that Bird 2 did not learn any category at all. Bird 2 seems to have learned the task the way pigeons learn a "pseudoconcept," where category membership is arbitrarily determined and not based on family resemblance. Bird 2 simply learned to associate a specific collection of strings with the left key and a different collection of strings with the right key. Consequently, its acquisition was slower, and it was not able to correctly categorize novel strings during transfer. On the other hand, Birds 1 and 4 relied on the family resemblance among grammatical strings in order to respond accurately to the novel strings presented during transfer.

Table 2 shows percentages of correct responses as a function of grammatical versus nongrammatical string and of string type and length, averaged over the 10 days of the transfer test that included novel strings. Performances on grammatical strings of Type 2 and on the shortest grammatical strings were examined to evaluate the possibility that strings of these types were memorized during training. If they were, performance on novel strings of those types would be predicted to be poor. This test, of course, applies only to Birds 1 and 4, who performed well on these strings during training. Table 2 shows that these birds did not perform particularly poorly on these strings during the novel string test. Their performance on strings of Type 2 was no longer the most accurate, but it was not as severely impaired as would have been expected if they had exclusively memorized Type 2 training strings. Accurate responses to Type 2 training strings apparently were not due exclusively to their rote memorization. Bird 1 continued to perform perfectly on the shortest novel strings, implying that that bird did not merely memorize the corresponding training strings. Bird 4, on the other hand, performed more poorly on the shortest novel strings than on any other type, implying that Bird 4 may have performed well on the shortest strings during training by having memorized them. Overall, some evidence for memorization of short stimuli was apparent, but evidence for having learned complex regularities was also apparent.

The bottom panel of Figure 2 shows that responding to training strings remained above chance during the transfer test. Birds 1 and 4 averaged 67% ($SD = 3.3$) accuracy on familiar strings (95% CI, 62%–71%; effect size $r = .991$). The most accurate responding continued to be to grammatical strings (83% correct, $SD = 6.4$%; 95% CI, 74%–92%; effect size $r = .991$), and responding to familiar nongrammatical strings remained indistinguishable from chance (51% correct responses; 95% CI, 50%–51%; effect size $r = .981$). In light of Bird 2's failure to transfer, it is interesting that it nevertheless continued in the transfer test to respond to familiar letter strings at greater than chance levels. Bird 2 classified 57% of grammatical and 65% of nongrammatical letter strings correctly, for an average of 61% accuracy across all the familiar strings. Thus, Bird 2's inability to correctly respond to novel strings was not due to a general breakdown of performance during the novel string transfer condition.

**Novel string transfer in terms of chunks**. The preceding analysis was in terms of accuracy of performance and was consistent with a position according to which AGL involved learning complex regularities. Overall accuracy, however, cannot directly address aspects of performance that can more decisively reveal whether pigeons learned complex regularities or memorized short chunks of training strings and then responded to novel strings on the basis of the chunks that they shared with training strings.

To discriminate between these two possibilities, we developed an analytical method that derived from the fact that the set of training strings was a subset of the complete set of strings in terms of which the grammar was defined and the regularities across the training strings only approximated the regularities of the grammar. For that reason, our definition of chunk strength did not take novel transfer strings into consideration: Chunk strength reflected only the degree to which reinforcement during training identified a chunk as correlated with the subset of grammatical strings used during training, and we emphasize that the subset did not perfectly reflect the regularities of the artificial grammar. Therefore, if, during training,
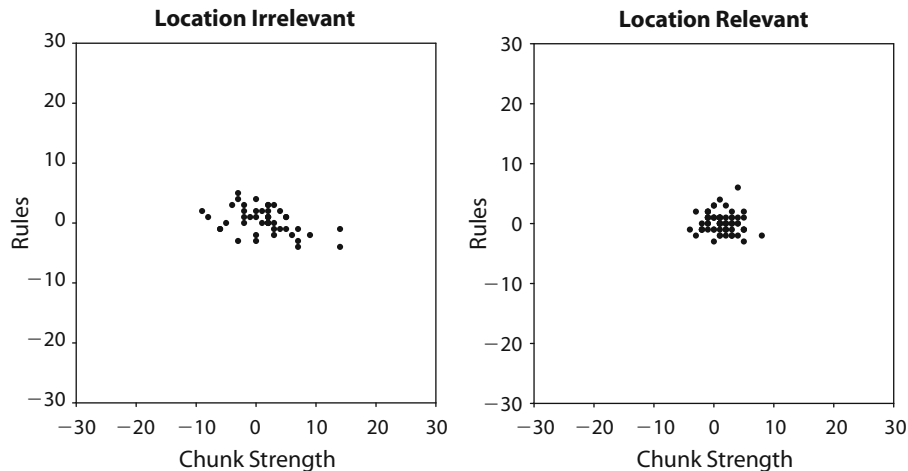
**Figure 3. Chunk strength of the transfer stimuli in Experiment 1 defined with respect to the grammar's rules (using the complete set of 86 stimuli, including those not seen during transfer), plotted as a function of the corresponding strength defined in terms only of the set of 62 training strings. The left panel does not take into account the location in which a chunk appeared, whereas the right panel does.**

a bird had memorized only short chunks correlated with reinforced strings, performance on novel transfer strings would be different than if it had learned more complex regularities as defined by the rules of the grammar shown at the bottom of Figure 1. Specifically, the pigeons would have responded to novel strings in terms of the strengths of their chunks defined with respect to training strings. Pigeons that had learned more complex sequential regularities, however, would have responded to novel transfer strings more in terms of their conformity to the regularities of the grammar. For simplicity, we will refer to these two different kinds of performances as *training-based* and *grammar-based* performances, respectively.

The logic and computational details of a new chunk-based analysis designed to distinguish between these two kinds of performances was as follows. Figure 3 pictures the logic by displaying optimal grammar-based performance as a function of optimal training-based performance. Each point in the left and right panels of Figure 3 represents a specific chunk occurring in any position within a novel string or in a specific location, respectively. Collectively, the points represent all the chunks in all the novel strings. The critical fact is that the two ways of disambiguating novel strings led to different responses, although only slightly so with location relevant. The correlations in the left and right panels of Figure 3—that is, the correlations between optimal training-based and optimal grammar-based responding—were −.43 and −.09 for location irrelevant and location relevant, respectively. The question was, then, whether performance was better described as training based or grammar based.

Figures 4 and 5 compare the two competing possibilities for Bird 1, which was representative in the sense that its mean accuracy to transfer stimuli was closest to the group mean. Figure 4 displays Bird 1's responses (number of left minus number of right responses) to novel transfer strings as a function of chunk strength defined with respect to the training strings. Figure 5 displays the same transfer performance as a function of the grammar. Perfect transfer based on the regularities in the grammar would result in points in Figure 5 lying directly on a diagonal with a slope of 10 and an intercept of 0. Thus, perfect transfer responding based on the reinforced training strings would produce a strong linear relation in Figure 4, and perfect responding based on the grammar would produce a strong linear relation in Figure 5. On the one hand, if responses to novel strings showed a stronger relation to optimal training-based performance than to optimal grammar-based performance, then, by definition, birds based responses more on training chunks than on the complex regularities in the grammar. On the other hand, if responses to novel strings showed a stronger relation to the grammar than to training, then, by definition, birds learned more about the regularities of the artificial grammar than just the information contained in the reinforcement histories of specific short chunks seen in the set of training strings.

How well did optimal training-based or optimal grammar-based performances describe the data? We need to ask only whether Birds 1 and 4 learned complex regularities, because, it will be recalled, Bird 2 failed to show such a tendency even by conventional standards. We therefore fit a linear regression line to the function shown for Bird 1 and another to the corresponding function for Bird 4 to determine whether performances of Birds 1 and 4 were better described as training based or grammar based. We then averaged the slopes and intercepts of the two regression lines, one for each of the two birds. Table 3 shows the average slope, average intercept, and average $r^2$. These averages were then compared using a separate 2 (strategy: chunk based or grammar based) $\times$ 2 (location: relevant or irrelevant) $\times$ 3 (chunk size: letter, bigram, or trigram) ANOVA for each of slope, intercept, and $r^2$. For slope, the main effect of strategy was significant at the .05 level [$F(1,12) = 5.19, p < .05$], and effect size $r$ was equal to a

**Figure 4. Representative Bird 1's responses (number of left responses minus number of right responses) to novel transfer strings in Experiment 1 as a function of chunk strength defined with respect to the set of 62 training strings. Panels represent responses to strings containing individual letters (top row), bigrams (middle row), and trigrams (bottom row) in specific locations (left column) or across all locations (right column).**
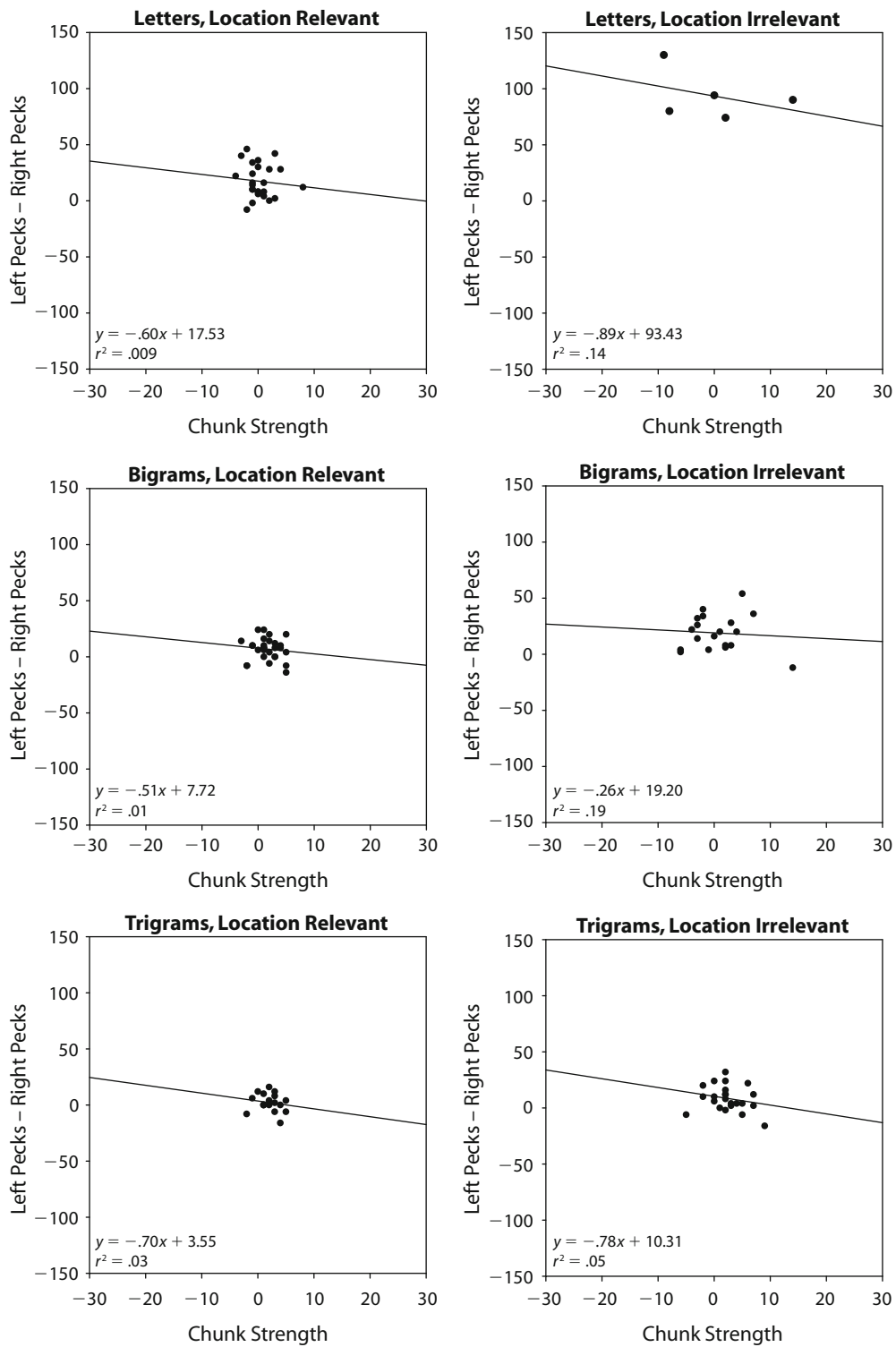
**Figure 5. Representative Bird 1's responses (number of left responses minus number of right responses) to novel transfer strings in Experiment 1 as a function of chunk strength defined with respect to the entire set of 86 possible strings produced by the grammar. Panels represent responses to strings containing individual letters (top row), bigrams (middle row), and trigrams (bottom row) in specific locations (left column) or across all locations (right column).**
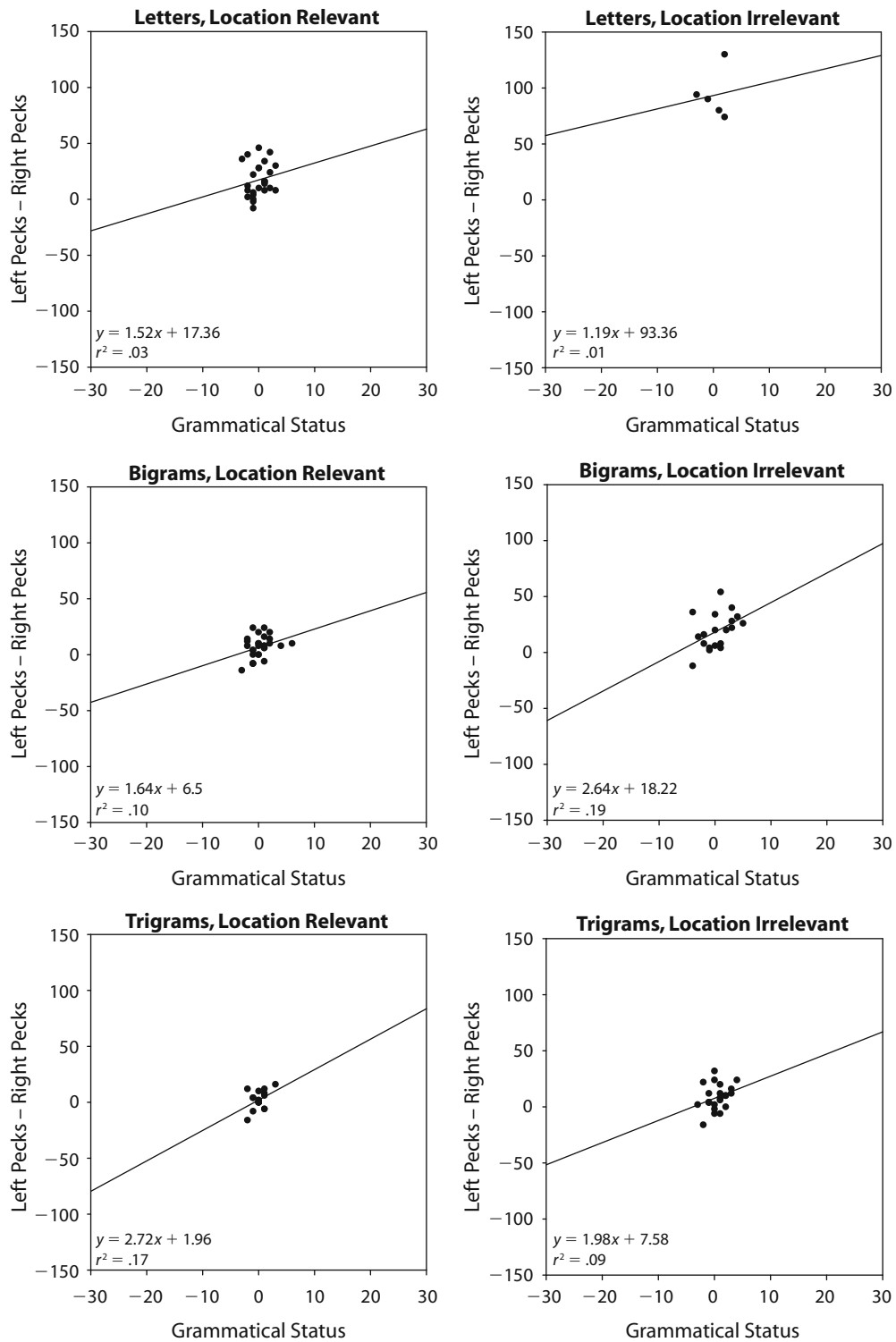
**Table 3**
**Coefficients of Determination ($r^2$), Slopes, and Intercepts for Responses**
**to Transfer Stimuli in Experiment 1**

| Stimuli | Location Relevant | | | Location Irrelevant | | |
|---|---|---|---|---|---|---|
| | $r^2$ | Slope | Intercept | $r^2$ | Slope | Intercept |
| | | | Chunks | | | |
| Letters | .05 ± .05 | 1.12 ± 0.74 | 16.895 ± 0.90 | .13 ± .02 | 1.07 ± 0.25 | 91.64 ± 2.53 |
| Bigrams | .01 ± .01 | −0.30 ± 0.30 | 9.07 ± 1.90 | .10 ± .13 | −0.10 ± 0.23 | 20.05 ± 1.20 |
| Trigrams | .02 ± .01 | −0.33 ± 0.53 | 4.17 ± 0.88 | .12 ± .09 | −0.92 ± 0.19 | 11.89 ± 2.23 |
| | | | Rules | | | |
| Letters | .18 ± .21 | 3.26 ± 2.46 | 17.04 ± 0.45 | .08 ± .09 | −2.54 ± 5.27 | 92.11 ± 1.77 |
| Bigrams | .33 ± .32 | 2.83 ± 1.68 | 7.91 ± 1.99 | .25 ± .08 | 3.50 ± 1.21 | 18.90 ± 0.95 |
| Trigrams | .12 ± .08 | 1.83 ± 1.27 | 3.40 ± 2.04 | .10 ± .01 | 1.97 ± 0.01 | 8.95 ± 1.94 |

Note—The optimal values are slope = 10 and intercept = 0.

medium .549. No other main effect or interaction was significant (all $F$s < 2.746, all $p$s > .104). For intercept, the main effects of chunk size [$F(2,12) = 1,838.93, p < .001$, effect size $r = .997$], and location relevance [$F(1,12) = 2,007.07, p < .001$, effect size $r = .997$] were significant, as was the interaction between chunk size and location relevance [$F(2,12) = 1,027.53, p < .001$, effect size $r = .994$]. No other main effects or interactions were significant (all $F$s < 1.715, all $p$s > .215). For $r^2$, all main effects and interactions were significant [chunk size, $F(2,12) = 32.31, p < .001$, effect size $r = .854$; location relevance, $F(1,12) = 34.53, p < .01$, effect size $r = .861$; strategy, $F(1,12) = 25.57, p < .001$, effect size $r = .825$; chunk size × location relevance, $F(2,12) = 34.29, p < .001$, effect size $r = .861$; chunk size × strategy, $F(2,12) = 39.49, p < .001$, effect size $r = .876$; location relevance × strategy, $F(1,12) = 28.10, p < .001$, effect size $r = .837$; and chunk size × location relevance × strategy, $F(2,12) = 36.69, p < .001$, effect size $r = .868$].

The critical findings were, first, in terms of slope, performance was described reliably better as grammar based than as training based, and the estimated slope based on regularities of the grammar more closely approximated optimal performance than did the estimated slope based on learning short chunks. Second, $r^2$ was reliably greater when performance was viewed in terms of regularities of the grammar than when it was viewed in terms of memorizing short chunks. A grammar-based analysis failed to describe intercepts reliably better than did a training-based analysis.

In summary, comparing training-based and grammar-based descriptions of performance showed that Birds 1 and 4 learned complex regularities: Performance of these birds could not be described purely in terms of learning short training chunks. Thus, both new and conventional analyses gave the same result: Birds 1 and 4 learned complex regularities of the grammar. Bird 2, however, responded in a qualitatively different way, gave no evidence of learning complex regularities, and seemed to memorize chunks and short strings.

## Discussion

Experiment 1 showed that performance of pigeons in an AGL task is controlled in part by complex sequential regularities produced by an artificial grammar. This demonstration rests on the criteria (1) that performance transferred to novel test stimuli and (2) that performance to these novel test stimuli was described better in terms of the regularities of the grammar than in terms of an assumption according to which pigeons learned short chunks of training strings and then, in transfer, responded on the basis of whether a test string contained those chunks. Experiment 1 showed also that pigeons memorized some easy strings and short chunks during training and then responded to novel transfer strings on the basis of chunks that they shared with training strings.

There were individual differences across pigeons in the extent to which complex regularities, as opposed to short stimuli, controlled categorization performance. Birds 1 and 4 were controlled more by complex regularities in the grammar than was Bird 2, who appeared to simply memorize short stimuli. Wright (2001) also found that given a great deal of training in a complex discrimination, some pigeons may memorize some stimuli. Compatible with this interpretation was that Bird 2 did not show the asymmetry shown by Birds 1 and 4, suggesting that the control acquired by specific stimuli applied equally to both grammatical and nongrammatical strings, as it would if Bird 2 simply memorized strings. Apart from Bird 2, Experiment 1 demonstrated that complex sequential regularities can control categorization performance of pigeons in an AGL task.

## EXPERIMENT 2

We hypothesized that there were two reasons why performance in Experiment 1 and in Herbranson and Shimp (2003) required extensive training and reached reliably better than chance but less than high levels of accuracy. First, performance in both experiments was asymmetric, in the sense that pigeons generally learned regularities of the grammar more accurately than distortions of those regularities. If we return to the plain English, intuitive idea of *family resemblance*, this is the result to be expected; members of a family visually resemble each other more than they resemble distortions of those individuals. We accordingly speculated that performance would be improved if pigeons were required only to learn regularities of grammars. Second, in Experiment 1 and in Herbranson and Shimp, each grammatical string had a nongrammatical counterpart with which it shared, by definition, all but a small number of chunks. This overlap of chunks between grammatical

and nongrammatical strings presumably increased the difficulty of the task by increasing the similarity between the two categories. We therefore again speculated that performance would be improved in a task in which pigeons had to learn only regularities of grammars.

We trained pigeons on two artificial grammars by presenting strings that were from either one grammar or another and by requiring the birds to peck different keys corresponding to different grammars. Strings from each grammar were presented randomly and equally often. One grammar was that used in the initial study by Herbranson and Shimp (2003; top of Figure 1) and the second was that used in Experiment 1 (bottom of Figure 1). The training task involved only grammatical strings, unlike the tasks in Herbranson and Shimp and Experiment 1, where pigeons were trained to learn grammatical and nongrammatical strings and then were tested with novel grammatical and novel nongrammatical strings.

We conducted two types of transfer tests, the first involving novel grammatical strings, and the second involving nongrammatical strings, which were also novel since no nongrammatical strings had been presented during training. These two tests were designed to evaluate two different hypotheses.

The first transfer test presented novel grammatical strings from each grammar and determined whether training would generalize to novel grammatical strings, much like the transfer test conducted in Experiment 1. It was, therefore, a test to determine whether pigeons simply memorized stimuli or learned complex regularities across exemplars in a manner consistent with the idea of family resemblance.

The second transfer test presented nongrammatical strings of each grammar. This transfer test allowed us to determine whether the birds would randomly respond to distortions of a grammar, respond to them as belonging to the grammar of which they were distortions, or respond to them as belonging to the other grammar. It will be recalled that in Experiment 1, the birds responded to distortions differently from the way they responded to grammatical strings. Here, however, we predicted the opposite result: We predicted that the difference between training and testing procedures in Experiments 1 and 2 would reverse how the pigeons responded to the very same novel nongrammatical strings. We predicted that a string that did not strictly satisfy one grammar would still be responded to as though it did, if the alternative was to respond to it as belonging to an entirely different grammar. Thus, we predicted that in Experiment 2, the pigeons would respond to nongrammatical strings the *same* way they responded to grammatical strings, even though, in Experiment 1, these very same nongrammatical strings had been responded to *differently* from grammatical strings. In short, the second transfer test evaluated the hypothesis that AGL in pigeons can be determined not only by the regularities inherent in a grammar, but also by the testing context.

## Method

**Animals and Apparatus**. Four experimentally naive birds participated. The apparatus was the same as that described in Experiment 1.

**Pretraining and Stimuli**. Pretraining and stimuli were the same as those in Experiment 1, with the one exception that the nongrammatical exemplars of a single grammar during training and transfer were replaced with grammatical exemplars generated from a second grammar. Appendix B shows all the strings used in Experiment 2.

**Procedure**. All 4 birds were given 32 days of training, by which time it was judged that all the birds had learned the task. Immediately after the end of this initial training, the birds were given 10 days of training, over which they were shown novel grammatical stimuli in a manner identical to that in Experiment 1. There were 24 novel grammatical stimuli, and each was presented once per day. After this first transfer test, the birds were given 15 days of training on the full set of possible grammatical strings, with all trials terminating in reinforcement. A subsequent 15-day transfer test involving string fragments taken out of context—specifically, with chunks presented in isolation—produced data that unsurprisingly, considering other data to be presented below that show how critical testing context is, were uninterpretable. These data are, therefore, not presented. Then the birds were given another 69 days of training with the full set of grammatical strings, and finally, the birds were given 40 days of training on a transfer test involving, for the first time in the experiment, nongrammatical strings. There were 88 nongrammatical strings, of which 22 were presented per day, so that over 4 days of training, each was presented once, and over the entire 40 days, each was presented 10 times. Nongrammatical strings were presented in a manner identical to that in the first novel transfer test. The strings themselves were the same as the nongrammatical strings in Experiment 1 and in Herbranson and Shimp (2003).

## Results

**Acquisition**. The 4 birds reached 60% overall accuracy between 22 and 36 days of training (mean, 26 days). This training period was only about one sixth that required in Herbranson and Shimp (2003) and in Experiment 1. It was so unexpectedly short that we chose to run each bird for at least 30 days—in part, to check that accuracy for the quickest-learning birds was not a statistical outlier. The top panel of Figure 6 shows that performance over the last 10 days of training averaged 74% ($SD = 9.35$) correct across all 4 birds (95% CI, 66–82; effect size $r = .946$). Accurate performance was, therefore, achieved much more quickly in Experiment 2 than in Experiment 1; accuracy reached a higher level, and the effect size was large.

A comparison between the top panel of Figure 6 and the corresponding top panel of Figure 2 for Experiment 1 shows that the asymmetry obtained in Herbranson and Shimp (2003) and in Experiment 1 was absent in Experiment 2: Bias for a key did not appear when, in Experiment 2, both keys represented a grammatical response. Responding to strings generated from Grammar 1 averaged 73% ($SD = 12.11$) correct (95% CI, 63%–84%; effect size $r = .911$), whereas responding to strings generated from Grammar 2 averaged 74% ($SD = 13.0$) correct (95% CI, 63%–85%; effect size $r = .906$).

The differences between Grammars 1 and 2 permitted a pigeon to discriminate between them by attending only to the first two or three letters. For example, if a string began with VT, it always belonged to Grammar 1 (bottom panel), or if it began with TP, it always belonged to Grammar 2 (top panel). Did the pigeons base their categorizations of strings only on the first two or three letters? If such had been the case, accuracy would not have depended on more complex structure involving letters after the initial two or
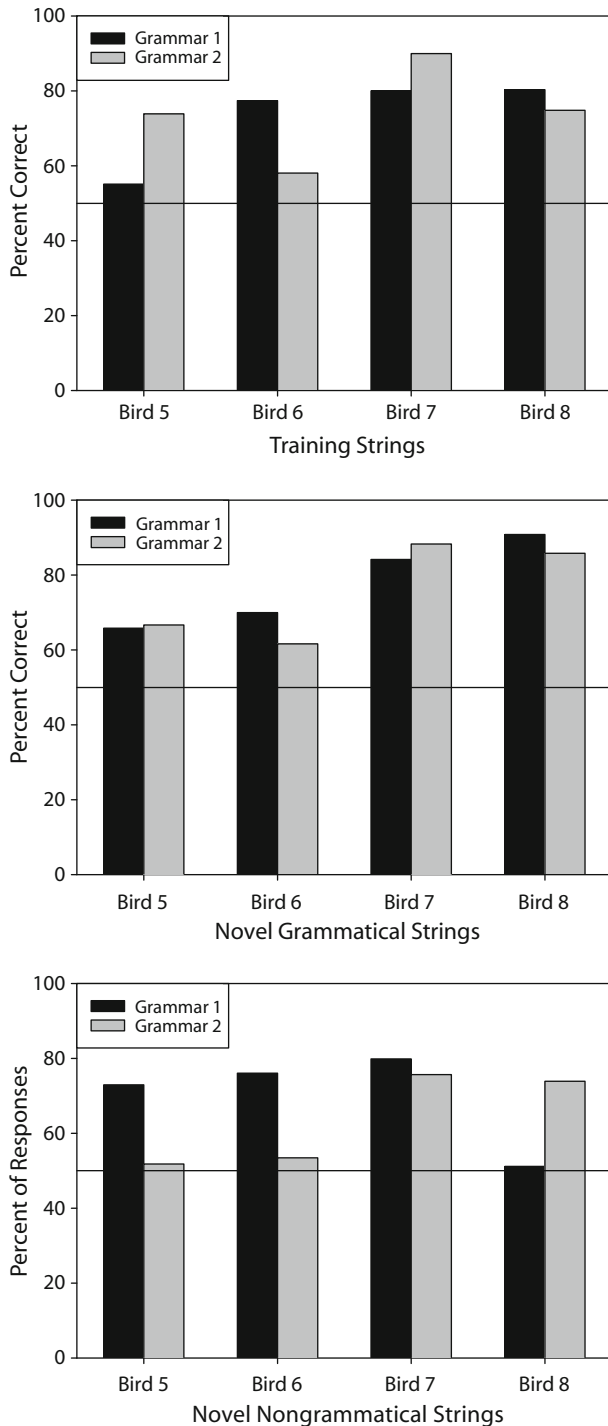
**Figure 6. Top: Percentages of correct responses to strings from Grammar 1 (dark bars) and strings from Grammar 2 (light bars) over the last 10 days of training for all 4 birds in Experiment 2. Middle: Percentages of correct responses to novel strings from Grammar 1 (dark bars) and novel strings from Grammar 2 (light bars) over 10 days of the first transfer for all 4 birds in Experiment 2. Bottom: Percentages of responses to nongrammatical distortions of strings produced by Grammar 1 (dark bars) and Grammar 2 (light bars) that corresponded to the grammar from which a string was distorted, over 40 days of the second transfer for all 4 birds in Experiment 2. In all three panels, the horizontal reference line indicates chance performance of 50%.**

three, and there would have been little or no variability in accuracy as a function of this more complex structure. Table 4 shows for each bird the means and standard deviations for accuracy across different strings having the same initial bigram or trigram. The column labeled "over different chunks" shows variability in how the birds responded to different initial chunks regardless of the rest of a string. Had the birds responded solely on the basis of initial chunks, accuracy would have been very high and would have shown little variability. Instead, accuracy was not close to 100%, and the standard deviation often overlapped chance (50%). The column labeled "over different strings" shows variability in how the birds responded to different strings that all began with the same initial chunk. If initial chunks were the basis for categorization, the remainder of a string would have little effect on responding, and there would have been little variability. Table 4 shows that, instead, there was considerable variability. In short, Table 4 shows that categorization depended on sequential structure more complex than initial bigrams or trigrams.

As in Experiment 1, the birds were required to learn grammatical strings of different string types and different string lengths. Table 5 shows the percentages of correct responses for each grammar as a function of string type and string length, averaged over the 4 birds and the last 10 days of training. For Grammar 1, string types were the same as those in Experiment 1 (bottom panel of Figure 1):

1. T(XS(X)P)V(P)S

2. TXT(P)S

3. V(S(X)PX)T(P)S

4. VS(X)PV(P)S

For Grammar 2, string types were as defined in Reber (1967) (see top of Figure 1):

1. T(P)TS

2. T(P)TX(X)(VPX(X))VS

3. T(P)TX(X)(VPX(X))VPS

4. V(X)(VPX(X))VS

5. V(X)(VPX(X))VPS

Table 5 shows few obvious tendencies for the birds to have been more accurate with some strings than with others, especially with what might seem intuitively easier types or lengths of strings. One such example appears to be that, for Grammar 2, grammatical strings of Type 2 were usually easier than strings of Type 1.

**First transfer test with novel strings: Novel grammatical strings**. The middle panel of Figure 6 shows that training transferred to 24 novel test strings. A comparison between the top and middle panels of Figure 6 shows that the birds responded to novel strings in much the same way as they did to training strings. Overall, 76% ($SD = 12.3$) of responses to these novel strings were categorized correctly (95% CI, 66%–87%; effect size $r = .929$).

Accuracy to novel strings during transfer was about equal for both grammars: 78% ($SD = 11.8$) of the strings gener-

**Table 4**
**Percentages of Correct Responses to Different Initial Bigrams and Trigrams (Over Different Chunks) and to Different Strings Beginning With the Same Initial Bigrams and Trigrams (Over Different Strings) in Experiment 2**

| | Bigrams | | Trigrams | |
|---|---|---|---|---|
| | Over Different Chunks | Over Different Strings | Over Different Chunks | Over Different Strings |
| Bird 5 | 63.2 ± 16.9 | 64.5 ± 21.2 | 63.8 ± 19.4 | 64.5 ± 21.2 |
| Bird 6 | 66.9 ± 16.1 | 66.6 ± 25.5 | 67.6 ± 17.0 | 66.6 ± 25.5 |
| Bird 7 | 84.6 ± 10.7 | 85.0 ± 15.1 | 85.1 ± 10.7 | 85.0 ± 15.1 |
| Bird 8 | 69.6 ± 28.0 | 77.7 ± 16.4 | 78.9 ± 11.3 | 77.7 ± 16.4 |

ated by Grammar 1 were correctly responded to (95% CI, 67%–88%; effect size $r = .939$), whereas 75% ($SD = 13.4$) of the strings generated by Grammar 2 were correctly responded to (95% CI, 64%–87%; effect size $r = .911$).

Birds continued in the transfer test to correctly respond to familiar strings, with 80% ($SD = 13.6$) correct for familiar strings (95% CI, 69%–92%; effect size $r = .931$). Responding to familiar strings was accurate for both grammars: Birds responded correctly to 80% ($SD = 14.8$) of strings from Grammar 1 (95% CI, 67%–92%; effect size $r = .918$) and 80% ($SD = 12.4\%$) of strings generated from Grammar 2 (95% CI, 70%–91%; effect size $r = .943$). Accuracy to familiar strings was somewhat better, but not reliably better [$t(3) = 2.215$, $p > .11$; effect size $r = .788$], during transfer than it was over the final 10 days of training, suggesting that the birds probably learned little more, if anything, during transfer and that performance had essentially reached a steady state when transfer began.

Table 6 shows percent correct for different string types and string lengths, averaged over the first transfer test. As was noted above, responding to strings of Type 2 in

Grammar 2 was accurate in training, and Table 6 shows that it continued to be accurate during transfer to novel exemplars of this type.

**Novel string transfer in terms of chunks**. We determined more directly, in the same manner as in Experiment 1, whether the birds learned complex regularities or just memorized chunks or strings. Figure 7 is the counterpart of Figure 3 and shows that in Experiment 2, as in Experiment 1, optimal training-based performance was correlated with, but not identical to, optimal grammar-based performance. Figure 7 shows also that training chunks were, on the whole, more diagnostic than chunks defined in terms of the grammar, in the sense that marginal variability was greater for training chunks than for the grammar-based chunks. Chunks defined in terms of training stimuli were much more diagnostic—that is, had greater marginal variability—than they were in Experiment 1 (Figure 3). The correlations in the left and right panels of Figure 7—that is, the correlations between optimal training-based performance and optimal grammar-based performance—were .80 and .75, respectively. Using Figure 7, we analyzed the data for all 4 birds in Experiment 2 in a manner analogous to that in Experiment 1. Figures 8 and 9 show the individual analysis for Bird 7, which was analogous to Bird 1 in Experiment 1, in the sense that its mean accuracy was closest to the group mean. Figure 8 shows representative Bird 7's transfer performance as a function of training-based chunks, and Figure 9 shows transfer performance as a function of grammar-based chunks. Visual inspection of the two figures suggests that grammar-based regularities may provide a better fit for the data than do training-based chunks.

Table 7 summarizes the results of this analysis for all 4 birds in Experiment 2. As in Experiment 1, we calculated 2 (strategy: chunk based or grammar based) × 2 (location: relevant or irrelevant) × 3 (chunk size: letter, bigram, or

**Table 5**
**Percentages of Correct Responses During Training in Experiment 2**

| | String Type | | | | | String Length | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 3 | 4 | 5 | 6 | 7 | 8 |
| **Bird 5** | | | | | | | | | | | |
| Grammar 1 | 66.0 | 74.0 | 77.8 | 80.0 | | 80.0 | 75.0 | 65.0 | 84.0 | 70.0 | 73.3 |
| Grammar 2 | 45.0 | 63.8 | 58.0 | 58.8 | 43.3 | 30.0 | 10.0 | 50.0 | 52.0 | 62.5 | 58.5 |
| Overall | | | | | | 55.0 | 53.3 | 56.0 | 68.0 | 66.5 | 65.6 |
| **Bird 6** | | | | | | | | | | | |
| Grammar 1 | 60.0 | 60.0 | 61.1 | 50.0 | | 70.0 | 60.0 | 80.0 | 58.0 | 60.0 | 51.7 |
| Grammar 2 | 32.5 | 96.3 | 86.0 | 82.5 | 68.3 | 80.0 | 60.0 | 86.7 | 74.0 | 73.8 | 80.0 |
| Overall | | | | | | 75.0 | 60.0 | 84.0 | 66.0 | 66.5 | 66.4 |
| **Bird 7** | | | | | | | | | | | |
| Grammar 1 | 93.0 | 92.0 | 90.0 | 84.3 | | 90.0 | 95.0 | 95.0 | 94.0 | 87.8 | 88.3 |
| Grammar 2 | 87.5 | 81.3 | 70.0 | 83.8 | 76.7 | 70.0 | 30.0 | 73.3 | 74.0 | 85.0 | 85.4 |
| Overall | | | | | | 80.0 | 73.3 | 82.0 | 84.0 | 86.5 | 86.8 |
| **Bird 8** | | | | | | | | | | | |
| Grammar 1 | 88.0 | 76.0 | 74.4 | 55.7 | | 90.0 | 80.0 | 85.0 | 90.0 | 66.7 | 70.8 |
| Grammar 2 | 75.0 | 88.8 | 82.0 | 81.3 | 70.0 | 80.0 | 60.0 | 80.0 | 74.0 | 77.5 | 86.2 |
| Overall | | | | | | 85.0 | 73.3 | 82.0 | 82.0 | 71.8 | 78.8 |
| **Average** | | | | | | | | | | | |
| Grammar 1 | 76.8 | 75.5 | 75.8 | 67.5 | | 82.5 | 77.5 | 81.3 | 81.5 | 71.1 | 71.0 |
| Grammar 2 | 60.0 | 82.5 | 74.0 | 76.6 | 64.6 | 65.0 | 40.0 | 72.5 | 68.5 | 74.7 | 77.5 |
| Overall | | | | | | 73.8 | 65.0 | 76.0 | 75.0 | 72.8 | 74.4 |

Note—There are no overall averages for string types, because string types are arbitrary labels. A given string type for one grammar has no special relationship to any string type for a different grammar.

**Table 6**
**Percentages of Correct Responses During First Transfer Test in Experiment 2**

| | String Type | | | | | String Length | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 3 | 4 | 5 | 6 | 7 | 8 |
| Bird 5 | | | | | | | | | | | |
| Grammar 1 | 75.0 | 46.7 | 80.0 | 67.5 | | 50.0 | 100.0 | 75.0 | 65.0 | 70.0 | 57.5 |
| Grammar 2 | 50.0 | 73.3 | 50.0 | 70.0 | 70.0 | 40.0 | 60.0 | 70.0 | 70.0 | 60.0 | 80.0 |
| Overall | | | | | | 45.0 | 73.3 | 73.3 | 67.5 | 64.0 | 67.1 |
| Bird 6 | | | | | | | | | | | |
| Grammar 1 | 60.0 | 56.7 | 66.7 | 62.5 | | 50.0 | 70.0 | 60.0 | 60.0 | 55.0 | 67.5 |
| Grammar 2 | 40.0 | 76.7 | 70.0 | 80.0 | 70.0 | 10.0 | 80.0 | 70.0 | 75.0 | 73.3 | 76.7 |
| Overall | | | | | | 30.0 | 76.7 | 63.3 | 67.5 | 66.0 | 71.4 |
| Bird 7 | | | | | | | | | | | |
| Grammar 1 | 85.0 | 96.7 | 90.0 | 82.5 | | 70.0 | 100.0 | 80.0 | 75.0 | 100.0 | 95.0 |
| Grammar 2 | 40.0 | 100.0 | 90.0 | 92.5 | 85.0 | 30.0 | 75.0 | 80.0 | 85.0 | 93.3 | 100.0 |
| Overall | | | | | | 50.0 | 83.3 | 80.0 | 80.0 | 96.0 | 97.1 |
| Bird 8 | | | | | | | | | | | |
| Grammar 1 | 85.0 | 96.7 | 96.7 | 70.0 | | 80.0 | 90.0 | 70.0 | 65.0 | 100.0 | 97.5 |
| Grammar 2 | 80.0 | 93.3 | 100.0 | 92.5 | 90.0 | 70.0 | 90.0 | 90.0 | 85.0 | 93.3 | 100.0 |
| Overall | | | | | | 75.0 | 90.0 | 76.7 | 75.0 | 96.0 | 98.6 |
| Average | | | | | | | | | | | |
| Grammar 1 | 76.3 | 74.2 | 83.3 | 70.6 | | 62.5 | 90.0 | 71.3 | 66.3 | 81.3 | 79.4 |
| Grammar 2 | 52.5 | 85.8 | 77.5 | 83.8 | 78.8 | 37.5 | 76.3 | 77.5 | 78.8 | 80.0 | 89.2 |
| Overall | | | | | | 50.0 | 80.8 | 73.3 | 72.5 | 80.5 | 83.6 |

Note—There are no overall averages for string types because string types are arbitrary labels. A given string type for one grammar has no special relationship to any string type for a different grammar.

trigram) ANOVAs on the resulting $r^2$, slope, and intercept. Analysis of $r^2$ produced significant main effects of strategy [$F(1,36) = 87.87$, $p < .001$, $r = .842$] and location [$F(1,36) = 16.25$, $p < .001$, $r = .558$], with the grammar-based and location-irrelevant analyses yielding better fits. No other main effects or interactions were significant (all $F$s $< 2.79$, all $p$s $> .07$). In addition, the analysis of slope produced a significant main effect of strategy [$F(1,36) = 62.16$, $p < .001$, $r = .796$], again with the grammar-based analysis providing a better fit. All of the remaining main effects and interactions were nonsignificant (all $F$s $< .488$, all $p$s $> .489$). The analysis of intercepts yielded

no significant main effects or interactions (all $F$s $< 2.45$, all $p$s $> .101$). The significant main effects of strategy provided evidence that the birds learned complex regularities: A grammar-based analysis gave a better overall description of the data than did a training-based analysis.

**Second transfer test with novel strings: Nongrammatical strings**. If the birds in Experiment 2 responded to novel nongrammatical strings strictly according to regularities conforming to grammatical structure, these strings would have failed, by definition, to satisfy the criteria for membership in either reinforced response class and the birds might, therefore, on the average, have responded
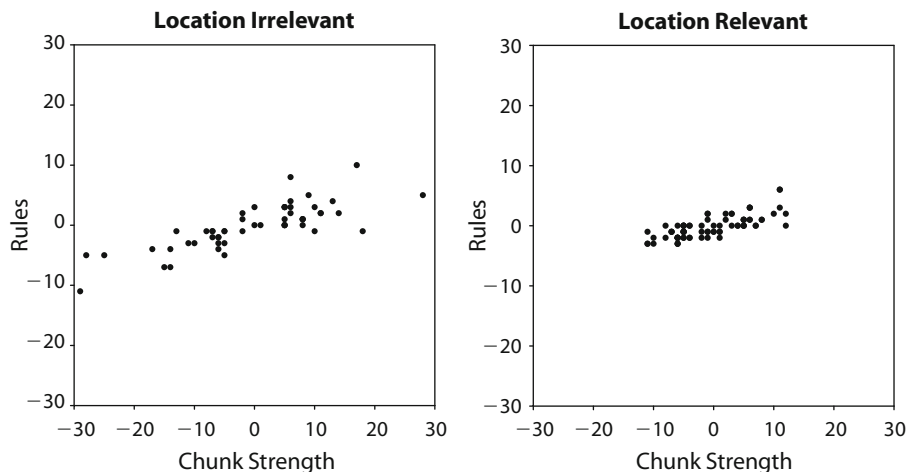


**Figure 7. Chunk strength of the grammatical transfer stimuli in Experiment 2 defined with respect to a grammar's rules (using the complete set of 86 stimuli, including those not seen during transfer) plotted against strength defined with respect to the corresponding grammar's set of 62 training strings. The left panel does not take into account the location in which a chunk appeared, whereas the right panel does.**
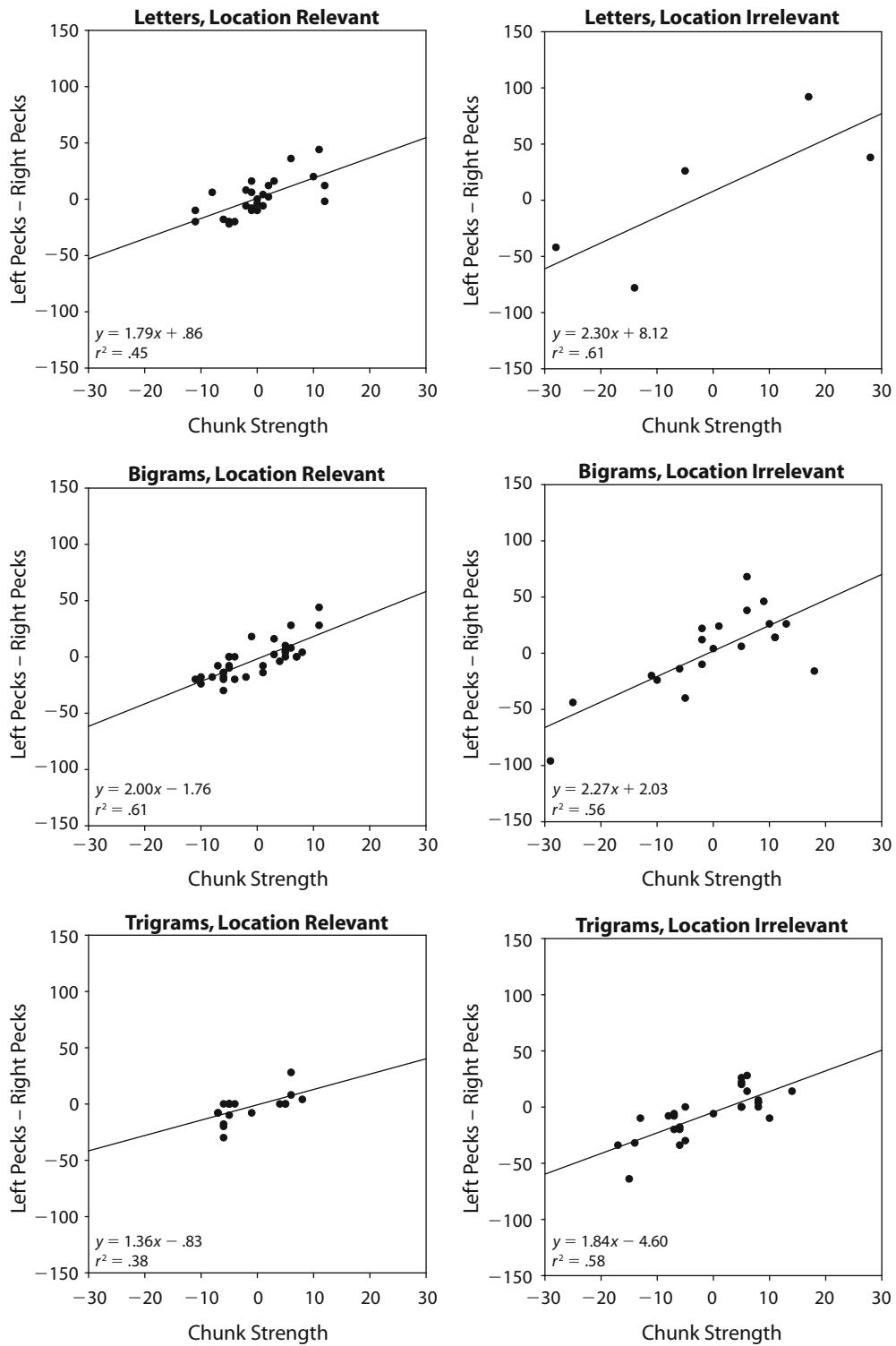
**Figure 8. Representative Bird 7's responses (number of left responses minus number of right responses) to novel transfer strings in Experiment 2 as a function of chunk strength defined with respect to the set of 62 training strings. Panels represent responses to strings containing individual letters (top row), bigrams (middle row), and trigrams (bottom row) in specific locations (left column) or across all locations (right column).**
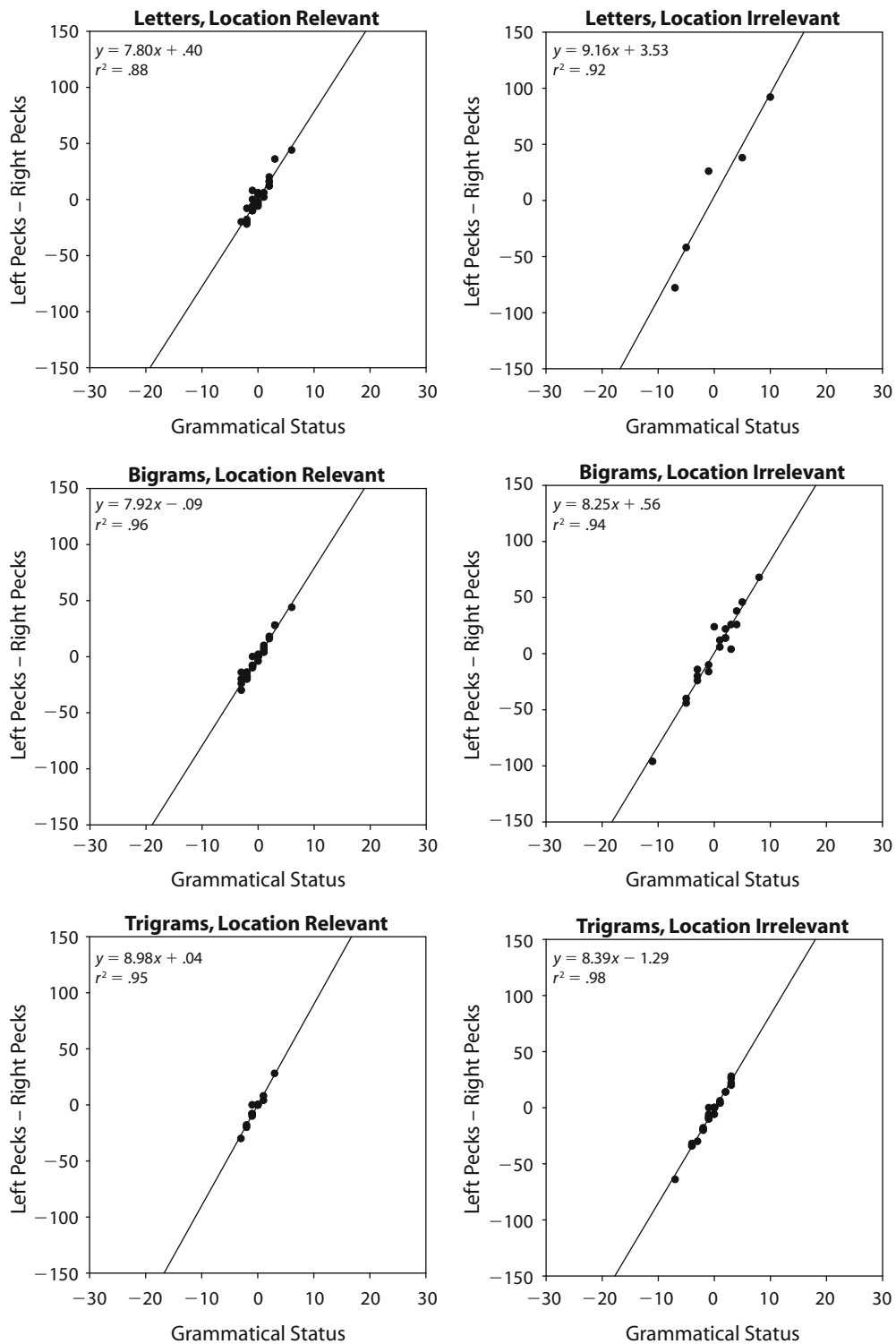
**Figure 9. Representative Bird 7's responses (number of left responses minus number of right responses) to novel transfer strings in Experiment 2 as a function of chunk strength defined with respect to the entire set of 86 possible strings produced by the grammar. Panels represent responses to strings containing individual letters (top row), bigrams (middle row), and trigrams (bottom row) in specific locations (left column) or across all locations (right column).**

**Table 7**
**Coefficients of Determination ($r^2$), Slopes, and Intercepts for Responses**
**to Transfer Stimuli in Experiment 2**

| Stimili | Location Relevant | | | Location Irrelevant | | |
|---|---|---|---|---|---|---|
| | $r^2$ | Slope | Intercept | $r^2$ | Slope | Intercept |
| | | | Chunks | | | |
| Letters | .39 ± .04 | 1.32 ± 0.25 | −1.99 ± 1.16 | .69 ± .07 | 1.78 ± 0.27 | 5.17 ± 8.62 |
| Bigrams | .56 ± .04 | 1.42 ± 0.35 | −0.74 ± 1.26 | .58 ± .02 | 1.69 ± 0.36 | −1.53 ± 1.54 |
| Trigrams | .33 ± .04 | 0.99 ± 0.27 | −0.76 ± 0.35 | .55 ± .04 | 1.35 ± 0.30 | −3.33 ± 0.97 |
| | | | Rules | | | |
| Letters | .73 ± .09 | 5.72 ± 1.14 | −2.33 ± 1.12 | .87 ± .03 | 6.63 ± 1.28 | −12.40 ± 7.07 |
| Bigrams | .84 ± .08 | 5.52 ± 1.42 | −1.08 ± 0.47 | .87 ± .04 | 5.89 ± 1.37 | −2.59 ± 1.38 |
| Trigrams | .80 ± .11 | 6.06 ± 1.76 | −0.80 ± 0.41 | .91 ± .04 | 6.06 ± 1.28 | −1.70 ± 0.48 |

Note—The optimal values are slope = 10 and intercept = 0.

randomly to them or, as pigeons sometimes do in such cases, might have adopted an extreme position bias. Alternatively, if the birds responded to distorted strings in terms of the degree to which those strings shared regularities with those from one or the other grammar, they would be predicted to respond to distorted strings as exemplars of the grammar of which they were distortions.

Table 8 shows mean performance over the second transfer test and shows that the birds were more likely to respond to distorted strings as belonging to the grammar of which they were garbled exemplars than as exemplars of the other grammar. (Table 8 does not show overall results for string type, because there were no string types in one grammar that were the same as those in the other grammar.)

The bottom panel of Figure 6 shows responses to distortions of strings from the two grammars during the second transfer test for all 4 birds. Dark bars represent the percentages of responses to distortions of strings from Grammar 1 that corresponded to the key previously associated with Grammar 1. Similarly, light bars represent the percentages of responses to distortions of strings from Grammar 2 that corresponded to the key previously associated with Grammar 2. The bottom panel of Figure 6 shows that when presented with a distortion of a string generated from Grammar 1, the pigeons pecked left 70.0% ($SD$ = 12.9) of the time, and when presented with a distortion of a string generated from Grammar 2, the pigeons pecked right 63.7% ($SD$ = 12.8) of the time. Thus, overall, birds responded to 66.3% ($SD$ = 7.8) of novel nongrammatical strings by pecking the key that produced reinforcement in the presence of undistorted grammatical strings generated from the same grammar. A two-tailed $t$ test confirmed that this tendency was reliable [$t(3)$ = 4.195, $p <$ .05; effect size $r$ = .943]. The pigeons discriminated between the two different sets of regularities correlated with the two different grammars and recognized distorted regularities of a grammar as more similar to the regularities of that grammar than to the regularities of the other grammar.

**Table 8**
**Percentages of Correct Responses to Nongrammatical Strings During**
**Second Transfer Test in Experiment 2**

| | String Type | | | | | String Length | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 3 | 4 | 5 | 6 | 7 | 8 |
| Bird 5 | | | | | | | | | | | |
| Grammar 1 | 69.0 | 70.7 | 71.4 | 81.9 | | 27.3 | 60.7 | 43.9 | 76.8 | 69.6 | 89.9 |
| Grammar 2 | 13.3 | 66.7 | 67.8 | 56.7 | 41.8 | 36.8 | 12.5 | 45.9 | 70.8 | 54.5 | 52.2 |
| Overall | | | | | | 31.7 | 35.0 | 44.9 | 73.8 | 62.2 | 70.0 |
| Bird 6 | | | | | | | | | | | |
| Grammar 1 | 72.6 | 74.4 | 69.5 | 89.8 | | 9.5 | 65.5 | 68.4 | 79.4 | 76.4 | 87.2 |
| Grammar 2 | 33.9 | 71.4 | 78.2 | 47.2 | 36.7 | 17.6 | 24.2 | 60.0 | 76.0 | 53.8 | 51.1 |
| Overall | | | | | | 15.8 | 43.5 | 64.4 | 77.7 | 65.3 | 68.2 |
| Bird 7 | | | | | | | | | | | |
| Grammar 1 | 80.2 | 73.2 | 79.0 | 85.7 | | 40.9 | 64.3 | 68.3 | 75.4 | 82.1 | 91.3 |
| Grammar 2 | 68.3 | 88.3 | 86.4 | 70.9 | 64.8 | 21.1 | 21.9 | 43.2 | 93.1 | 83.6 | 86.0 |
| Overall | | | | | | 31.7 | 41.7 | 56.4 | 84.4 | 92.9 | 88.5 |
| Bird 8 | | | | | | | | | | | |
| Grammar 1 | 58.0 | 40.8 | 47.3 | 56.0 | | 5.3 | 37.9 | 29.4 | 46.0 | 62.4 | 59.0 |
| Grammar 2 | 58.0 | 87.4 | 98.1 | 71.8 | 54.5 | 6.3 | 48.3 | 91.2 | 89.2 | 73.2 | 75.3 |
| Overall | | | | | | 5.9 | 43.1 | 60.3 | 67.8 | 67.7 | 67.6 |
| Average | | | | | | | | | | | |
| Grammar 1 | 70.0 | 64.8 | 66.8 | 78.4 | | 20.8 | 57.1 | 52.5 | 69.4 | 72.6 | 81.9 |
| Grammar 2 | 43.4 | 78.5 | 82.6 | 61.7 | 49.5 | 20.5 | 26.7 | 60.1 | 82.3 | 66.3 | 66.2 |
| Overall | | | | | | 20.7 | 41.9 | 56.3 | 75.9 | 69.5 | 74.1 |

Note—There are no overall averages for string types because string types are arbitrary labels. A given string type for one grammar has no special relationship to any string type for a different grammar.

**Table 9**
**Regression Analysis for Experiment 1**

|  | Training | | | Transfer | | |
|---|---|---|---|---|---|---|
|  | Nongrammatical | Grammatical | All | Nongrammatical | Grammatical | All |
| Bird 1 | 51.9 | 71.6 | 61.8 | 38.3 | 72.5 | 55.4 |
| Letters only | 45.4 | 65.1 | 55.3 | 40.5 | 58.3 | 49.4 |
| Bigrams only | 52.4 | 72.1 | 62.3 | 39.3 | 55.6 | 47.4 |
| Trigrams only | 51.9 | 71.6 | 61.8 | 43.5 | 52.4 | 48.0 |
| All chunks | 52.0 | 71.5 | 61.8 | 37.9 | 55.1 | 46.5 |
| Bird 2 | 69.7 | 58.7 | 64.2 | 45.0 | 47.5 | 46.3 |
| Letters only | 62.7 | 51.5 | 57.1 | 60.0 | 39.5 | 49.8 |
| Bigrams only | 68.5 | 57.6 | 63.1 | 62.8 | 43.8 | 53.3 |
| Trigrams only | 69.7 | 58.7 | 64.2 | 55.9 | 51.9 | 53.9 |
| All chunks | 81.6 | 58.5 | 70.0 | 44.3 | 55.9 | 50.1 |
| Bird 4 | 54.8 | 86.8 | 70.8 | 51.7 | 77.5 | 64.6 |
| Letters only | 40.3 | 70.9 | 55.6 | 36.1 | 66.4 | 51.2 |
| Bigrams only | 53.9 | 84.3 | 69.1 | 52.9 | 66.6 | 59.7 |
| Trigrams only | 56.1 | 86.8 | 71.5 | 58.3 | 66.6 | 62.4 |
| All chunks | 56.2 | 86.8 | 71.5 | 60.4 | 59.1 | 59.8 |
| Optimal |  |  |  |  |  |  |
| Letters only | 58.0 | 58.5 | 58.3 | 49.3 | 48.3 | 48.8 |
| Bigrams only | 78.8 | 83.2 | 80.1 | 68.9 | 53.8 | 61.4 |
| Trigrams only | 100.0 | 100.0 | 100.0 | 70.8 | 91.7 | 81.3 |
| All chunks | 100.0 | 100.0 | 100.0 | 75.0 | 100.0 | 87.5 |

**An alternative analysis of Experiments 1 and 2**. The chunk-based analyses of the results of Experiments 1 and 2 presented above involved differences between frequencies of chunks in training strings and in testing strings. In this sense, these analyses weighted more frequent chunks more heavily than less frequent chunks. In order to determine whether this asymmetry might have influenced our interpretation, we developed still another new analysis, one that did not weight frequent chunks more heavily. This method worked as follows. Multiple regression was used to predict a bird's responses to transfer stimuli from information about the reinforcement histories of specific chunks in those stimuli. We represented each letter string in terms of 155 predictor variables. Each variable corresponded to one of the possible chunks that could appear in the string (5 letters, 25 bigrams, and 125 trigrams) and assumed a value of either 0 or 1, according to whether the corresponding chunk was present or absent. For example, the letter string VTS had values of 1 for chunks consisting of V, T, S, VT, TS, and VTS. Each of the variables representing the remaining 149 chunks had values of 0.

Associated with each string was a variable representing the percentage of times a bird categorized it as grammatical over the last 10 days of training, ranging from 0 (always categorized as nongrammatical) to 1 (always categorized as grammatical). For each bird, we calculated a regression line that predicted this variable from the 155 chunk variables. On the one hand, if a bird learned to respond to strings only on the basis of their component chunks seen during training, this regression line should predict the bird's performance during both training and transfer. On the other hand, if a bird learned complex regularities beyond those captured by specific chunks, transfer performance might be better than predicted by the regression line. The results of this analysis are shown in Tables 9 and 10 for Experiments 1 and 2, respectively.

Tables 9 and 10 show that predictive accuracy for each bird was poorest when based on five single letters, better when based on 25 bigrams, still better based on 125 trigrams, and best when based on the full set of 155 chunks. This result simply reflects the fact that chunk-based regression estimates were more accurate when based on more chunk information. The regression line based on all 155 chunks closely approximated each bird's training performance, and there was no statistically reliable difference between the performance of birds over the final 10 days of training and their corresponding regression lines [$t(6) = 0.964, p > .05$, effect size $r$ equal to a small .366].

There *was* a reliable difference, however, between the birds' performance to novel *transfer* strings and predictions made by the corresponding regression lines. The birds reliably outperformed the regression lines [$t(6) = 2.769, p < .05$, effect size $r$ equal to a medium .749, where .800 defines a large effect size]. Every bird in both experiments conformed to this pattern, except for one individual case, Bird 2 from Experiment 1, who, it will be recalled, responded chiefly on the basis of memorizing specific stimuli. If this one instance is excluded from the analysis on the grounds that Bird 2 clearly did not learn abstract structure, the difference was greater still [$t(5) = 5.160, p < .005$, effect size $r$ equal to a large .918].

In addition to the summaries for each bird, Tables 9 and 10 also summarize the performance of a hypothetical optimal bird. These optimal values were calculated in the same way as for real birds, but on the basis of the performance of a hypothetical bird that achieved 100% accuracy during training—that is, that always pecked left to grammatical strings in Experiment 1 (Table 9) and to strings from Grammar 1 in Experiment 2 (Table 10) and, correspondingly, always pecked right to nongrammatical strings in Experiment 1 (Table 9) and to strings from Grammar 2 in Experiment 2 (Table 10). These values show

**Table 10**
**Regression Analysis for Experiment 2**

|  | Training | | | Transfer | | |
|---|---|---|---|---|---|---|
|  | Grammar 1 | Grammar 2 | All | Grammar 1 | Grammar 2 | All |
| Bird 5 | 55.2 | 73.9 | 64.5 | 65.8 | 66.7 | 66.3 |
| Letters only | 51.5 | 71.7 | 61.6 | 47.0 | 74.5 | 60.8 |
| Bigrams only | 54.0 | 73.7 | 63.8 | 49.6 | 82.4 | 66.0 |
| Trigrams only | 53.9 | 73.8 | 63.9 | 45.0 | 81.2 | 63.1 |
| All chunks | 53.7 | 73.9 | 63.8 | 35.0 | 84.0 | 59.5 |
| Bird 6 | 77.4 | 58.1 | 67.7 | 70.0 | 61.7 | 65.8 |
| Letters only | 73.1 | 54.1 | 63.6 | 69.4 | 52.4 | 60.9 |
| Bigrams only | 77.3 | 59.0 | 68.2 | 70.8 | 60.1 | 65.5 |
| Trigrams only | 77.2 | 58.0 | 67.6 | 75.3 | 52.2 | 63.8 |
| All chunks | 77.0 | 58.1 | 67.6 | 76.8 | 52.5 | 64.7 |
| Bird 7 | 80.0 | 90.0 | 85.0 | 84.2 | 88.3 | 86.3 |
| Letters only | 67.9 | 77.7 | 72.8 | 62.9 | 86.1 | 74.5 |
| Bigrams only | 80.1 | 89.9 | 85.0 | 75.7 | 95.4 | 85.6 |
| Trigrams only | 79.9 | 90.0 | 85.0 | 78.6 | 87.5 | 83.0 |
| All chunks | 79.9 | 90.0 | 85.0 | 65.7 | 89.7 | 77.7 |
| Bird 8 | 80.3 | 74.8 | 77.6 | 90.8 | 85.8 | 88.3 |
| Letters only | 70.1 | 64.5 | 67.3 | 66.9 | 70.4 | 68.6 |
| Bigrams only | 80.3 | 74.8 | 77.5 | 78.3 | 75.5 | 76.9 |
| Trigrams only | 80.3 | 74.6 | 77.5 | 79.0 | 73.1 | 76.0 |
| All chunks | 80.3 | 74.9 | 77.6 | 81.9 | 77.2 | 79.5 |
| Optimal |  |  |  |  |  |  |
| Letters only | 79.8 | 80.5 | 80.1 | 76.8 | 88.7 | 82.8 |
| Bigrams only | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Trigrams only | 100.0 | 100.0 | 100.0 | 100.0 | 91.7 | 95.8 |
| All chunks | 100.0 | 100.0 | 100.0 | 100.0 | 91.7 | 95.8 |

that performance based on the optimal use of chunks contained in the set of training strings would sometimes have been as accurate or even more accurate than that of our pigeons, and this optimal chunk-based performance would not require learning any regularities more complex than those in trigrams. This analysis would seem to support a chunk-based interpretation of our results, except that Tables 9 and 10 also show that predictions made by the regression lines did not describe the crucial results from transfer tests. Bird 8 in Experiment 2, for example, displayed an overall transfer accuracy (88.3%) greater than that predicted by the chunk-based regression line (79.5%) and far less than optimal chunk-based accuracy (95.8%). The remaining birds (again with the exception of Bird 2) showed the same pattern. That is, chunk-based predictions could accurately predict training performances, but not birds' transfer performances.

Overall, the birds in Experiment 2 learned to a high level of accuracy, accurately transferred performance to novel stimuli, did not simply discriminate on the basis of initial bigrams or trigrams, and showed by new criteria that they learned two sets of complex regularities correlated with two different grammars.

## Discussion

We speculated on the basis of an intuitive, plain English interpretation of *family resemblance* that performance would be acquired much more quickly and more accurately in Experiment 2 than in Experiment 1 and in Herbranson and Shimp (2003), and it was. We further speculated that in Experiment 2, the preference for one response key observed in Experiment 1 and in Herbranson

and Shimp would be eliminated, and it was. Experiment 2 therefore supports the idea that pigeons can quickly and accurately learn complex sequential regularities across strings generated by artificial grammars. At the same time, there was evidence to suggest that pigeons also memorized some easy strings and short chunks. Therefore, as in Experiment 1 and in Herbranson and Shimp, pigeon AGL involved both memorizing some stimuli and learning complex regularities.

We believe that learning in Experiment 2 was superior because its two stimulus categories were less similar to each other than was the case in Experiment 1 and in Herbranson and Shimp (2003). It should be noted that the differences among the correlations in Figures 3 and 7 depended on the relation between training and transfer strings, so they could not have had anything to do with the original rate of learning during training. In short, the results are compatible with the idea that within each category in Experiment 2, the strings shared a family resemblance and that the two categories shared less of a family resemblance to each other than they did in Experiment 1.

## GENERAL DISCUSSION

### Pigeon AGL Involves Learning Complex Sequential Regularities

Experiments on human AGL have used sequences of either auditory or visual stimuli and have shown that performance can involve both memorizing stimuli and learning complex regularities across stimuli (Gomez & Gerken, 1999; Knowlton & Squire, 1996; McAndrews & Moscovitch, 1985; Perruchet & Pacteau, 1990). These complex

sequential regularities can control performance without a person's being able to explicitly describe any rules or even being aware of having learned regularities, thus making it more plausible that nonverbal nonhuman animals might be able to engage in an analogous form of learning. We therefore asked whether pigeons could demonstrate AGL and, if so, whether they did so simply by memorizing chunks of strings or by learning both chunks and more complex regularities across strings. We found evidence that pigeons both memorized chunks and learned more complex sequential regularities. These results agree with and extend those we originally obtained (Herbranson & Shimp, 2003): They were obtained with a new artificial grammar, and several new statistical analyses give converging support for our original results. These analyses provide new evidence that pigeons acquired information about regularities more complex than that obtainable from short chunks as conventionally defined in AGL tasks. We also determined that the pigeons in Experiment 2 did not categorize strings simply on the basis of the initial few letters (see Table 4).

We believe that these various analyses provide powerful evidence in favor of the view that pigeon AGL involves learning complex regularities. We also found, however, that pigeon AGL involves memorizing some short patterns of letters as well. Experiment 1, for example, showed that performances by Birds 1 and 4 were controlled by both complex regularities and simple stimuli, such as strings of length 3 and strings of Type 1. This result reminds us of both Wright's (2001) claim that pigeons either memorized specific stimuli or learned rules in matching-to-sample tasks and Loidolt, Aust, Meran, and Huber's (2003) claim that pigeons can use either specific or general information. Memorization of some stimuli is to be expected, because the birds were given extensive training and because some training strings, especially short strings and easy strings, were few enough in number for the pigeons to memorize (Wright, Cook, Rivera, Sands, & Delius, 1988). These pigeon data, as well as our own, correspond in this way to a position in the human literature according to which AGL involves both memorizing stimuli and learning complex regularities (Knowlton & Squire, 1996; Perruchet & Pacteau, 1990).

Finally, we found a critical role for the testing context in determining what pigeons learn in AGL tasks. The second transfer test of Experiment 2 showed that strings from the same artificial grammar were responded to differently, depending on the testing context. Thus, the regularities implicit in strings generated by an artificial grammar do not by themselves determine how pigeons respond to those strings. Future theories of AGL in pigeons will have to take into account the relation between training and testing contexts, not simply the regularities implicit in an artificial grammar.

## Qualifications

First, the very term *AGL* must be used with delicacy when the subjects are birds. Although the similarities between avian AGL tasks and the conventional human version are sufficient in number to justify the same label for both, there are important differences. In nearly all human AGL tasks, during training, subjects are shown only grammatical strings (see, however, Dienes, Broadbent, &

Berry, 1991; Gomez & Lakusta, 2004), whereas, in contrast, the pigeons in Herbranson and Shimp (2003) and in both of the present experiments were reinforced during training for correct responses to two alternative types of strings and Gentner et al. (2006) used an analogous task with starlings. Such procedural differences might make a bird task qualitatively different from a human task. Second, pigeons apparently learn regularities related to those of an artificial grammar, but there is no justification for an extreme claim that the regularities they learn are either exactly those of the grammar itself or whatever regularities humans learn. Third, we emphasize that we make no claim that the demonstration that pigeons can learn complex regularities related to those of an artificial grammar necessarily implies that they can learn anything like the grammar of a natural language, although the possibility remains open that there may be some sense in which they can do so.

## Conceptual Interpretations of Complex Stimulus Control

There is no dearth of ideas to explain how nonhuman animals, especially pigeons, learn to discriminate among complex visual categories, including naturalistic and ill-defined categories that have been interpreted in terms of family resemblance. We believe that several of these ideas show promise and that our data identify important new goals for their development. A better scientific understanding of the intuitive, plain English term *family resemblance* could be especially fruitful because, having been imported into psychology in the first place from conceptual analyses of visual perception and grammar (Wittgenstein, 1953), it has already been subjected to an unusually high degree of conceptual and linguistic scrutiny. One promising idea is multidimensional scaling based on similarity judgments. This approach has been applied to categorizations by monkeys of naturalistic stimuli, such as human faces and fruit (Sands, Lincoln, & Wright, 1982). This approach nicely handles spatial regularities inherent in faces and fruit. Our data highlight the importance of generalizing this approach to handle the spatial regularities in sequences of letters of an artificial grammar. We believe that such a generalization would provide an unusually powerful and general understanding of how pigeons learn spatial regularities. Our own work on optimal categorization by pigeons of rectangles varying in height and width or moving objects varying in speed and direction could also profit from a similar generalization. Another perspective on pigeon visual categorization of complex stimuli is that of prototype theory, in which a set of possible exemplars is represented by its average. We are less sanguine about the future promise of this approach, because our data make us wonder what the average grammatical string would look like and what its function in categorization could possibly be. Lastly, and in our judgment perhaps the most promising idea, is that of dynamic interacting systems, as illustrated by connectionist models that have provided useful interpretations of many complex perceptual phenomena. Vokey and Tangen (2001) used a connectionist model to simulate categorization in an ex-

periment by Watanabe, Sakamoto, and Wakita (1995), in which pigeons learned to discriminate between paintings in different styles (e.g., Picasso vs. Monet). Their simulations generated eight eigenvectors (see Figure 7 in Vokey & Tangen, 2001) that they suggest may have been a basis for the pigeons' discriminations of artistic style. Our data could facilitate development of connectionist models for complex discrimination learning because it would be especially important, we believe, if connectionist models could provide estimates of eigenvectors for discriminating grammatical from nongrammatical strings of letters and show how those eigenvectors are related to those for discriminating artistic style. If these different tasks generated eigenvectors that could somehow be theoretically connected or explained, a general connectionist model might emerge to handle both naturalistic images and grammatical regularities, and a most impressive unification of Wittgenstein's idea of family resemblance might result.

## Comparative Perspectives

We would be remiss if we failed to comment on the broader issue of the similarities between natural language in humans and language-like behaviors in nonhumans. Research on this issue has focused primarily on monkeys and nonhuman primates (e.g., Gardner & Gardner, 1978; Hauser, Newport, & Aslin, 2001; Premack, 1976; Rumbaugh & Washburn, 2003), and theories of the origins of language and of language development have emphasized the behavioral, cognitive, ecological, and neuroanatomical similarities between humans and primates (Greenfield, 1991; Hauser et al., 2002; Krasnegor, Lyon, & Goldman-Rakic, 1997; Lieberman, 2002).

This sharp focus on the similarity of linguistic abilities and neuroanatomical structures across similar species has proven merit but may not facilitate our seeing continuity of component processes of natural language across less similar species. Fitch, Hauser, and Chomsky (2005) suggested that comparative research on these component processes with diverse species should have high priority. Similarly, Emery and Clayton (2004) suggested that apes and crows have "cognitive tool kits" with some tools in common. They suggested that cognition in apes and corvids may have evolved through a process of convergent mental evolution but divergent brain evolution. The present results suggest that pigeons have a capacity for learning complex regularities in systems of spatial strings of arbitrary visual alphabetic stimuli. Combined with the fact that starlings can recognize regularities across temporal sequences of their naturalistic perceptual auditory units, our results suggest a general ability of birds to learn to recognize regularities in systems of complex sequential stimuli. We believe the possibility of this general ability is further strengthened by the discovery that pigeons can learn to recognize statistical regularities in repeating temporal sequences of spatial locations and do so in a manner highly similar to that of humans (Froehlich, Herbranson, Loper, Wood, & Shimp, 2004; Shimp, Froehlich, & Herbranson, 2007). This cognitive tool may be a prerequisite for human language but now appears not to be uniquely human or even uniquely mammalian.

## 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.

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

DIENES, J. A., BROADBENT, D., & BERRY, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 875-887.

EMERY, N. J., & CLAYTON, N. S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science*, **306**, 1903-1907.

FITCH, W. T., HAUSER, M. D., & CHOMSKY, N. (2005). The evolution of the language facility: Clarifications and implications. *Cognition*, **97**, 179-210.

FROEHLICH, A. L., HERBRANSON, W. T., LOPER, J. D., WOOD, D. M., & SHIMP, C. P. (2004). Anticipating by pigeons depends on local statistical information in a serial response time task. *Journal of Experimental Psychology: General*, **133**, 31-45.

GARDNER, R. A., & GARDNER, B. T. (1978). Comparative psychology and language acquisition. In K. Salzinger & F. L. Denmark (Eds.), *Psychology: The state of the art* (Annals of the New York Academy of Sciences, Vol. 309, pp. 37-76). New York: New York Academy of Sciences.

GENTNER, T. Q., FENN, K. M., MARGOLIASH, D., & NUSBAUM, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, **440**, 1204-1207.

GOMEZ, R. L., & GERKEN, L. A. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition*, **70**, 109-135.

GOMEZ, R. L., & LAKUSTA, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, **7**, 567-580.

GREENFIELD, P. M. (1991). Language, tools, and the brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral & Brain Sciences*, **14**, 531-551.

HAUSER, M. D., CHOMSKY, N., & FITCH, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, **298**, 1569-1579.

HAUSER, M. D., NEWPORT, E. L., & ASLIN, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, **78**, 53-64.

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., & SHIMP, C. P. (2003). "Artificial grammar learning" in pigeons: A preliminary analysis. *Learning & Behavior*, **31**, 98-106.

HERRNSTEIN, R. J., LOVELAND, D. H., & CABLE, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **2**, 285-302.

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.

KRASNEGOR, N. A., LYON, O. R., & GOLDMAN-RAKIC, P. S. (1997).

*Development of the prefrontal cortex: Evolution, neurobiology, and behavior*. Baltimore: Brookes.

LIEBERMAN, P. (2002). On the nature and evolution of the neural bases of human language. In C. Ruff (Ed.), *Yearbook of physical anthropology* (pp. 36-62). New York: Wiley-Liss.

LOIDOLT, M., AUST, U., MERAN, I., & HUBER, L. (2003). Pigeons use item-specific and category-level information in the identification and categorization of human faces. *Journal of Experimental Psychology: Animal Behavior Processes*, **29**, 261-276.

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.

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

PREMACK, D. (1976). *Intelligence in ape and man*. Hillsdale, NJ: Erlbaum.

REBER, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior*, **6**, 855-863.

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

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.

ROSCH, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

ROSNOW, R. L., & ROSENTHAL, R. (2003). Effect sizes for experimental psychologists. *Canadian Journal of Experimental Psychology*, **57**, 221-237.

RUMBAUGH, D. M., & WASHBURN, D. A. (2003). *Intelligence of apes and other rational beings*. New Haven, CT: Yale University Press.

SANDS, S. F., LINCOLN, C. E., & WRIGHT, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. *Journal of Experimental Psychology: General*, **111**, 369-389.

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., FROEHLICH, A. L., & HERBRANSON, W. T. (2007). Information processing in pigeons: Incentive as information. *Journal of Comparative Psychology*, **121**, 73-81.

VOKEY, J. R., & TANGEN, J. M. (2001, June). *Learning an artist's style: Just what does a pigeon see in a Picasso?* Paper presented at the annual meeting of the Canadian Society for Brain, Behavior & Cognitive Science, Quebec. Available at people.uleth.ca/~vokey/pdf/WatanabeSimposter .pdf.

WATANABE, S., SAKAMOTO, J., & WAKITA, M. (1995). Pigeons' discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior*, **63**, 165-174.

WITTGENSTEIN, L. (1953). *Philosophical investigations*. New York: Macmillan.

WRIGHT, A. A. (2001). Learning strategies in matching to sample. In R. G. Cook (Ed.), *Avian visual cognition* [Online]. 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.

## APPENDIX A
### Experiment 1 Stimuli

Training Stimuli

| Grammatical | | Nongrammatical | |
| --- | --- | --- | --- |
| VTS | TXTS | VVS | TXPS |
| VTPS | TVPPS | VXSPVS | TSPXTS |
| TXTPS | TVPPPS | VSSVPS | TPSXPVS |
| TXSPVS | TXTPPS | TXSPXSS | VSXSPVS |
| VSPXTS | VTPPPS | VSPTPPS | VSXSPXTVS |
| TVPPPPS | TXSPVPS | TXSSXPVS | TXSPXTVS |
| TXSXPVS | TXTPPPS | VSXVXTPS | VTPSPPPS |
| VSXPXTS | VTPPPPS | VXPXTPPS | TSXPVPPS |
| VSPVPPS | VSXPVPS | TPVS | TTPSS |
| VSXXPVS | TVPPPPPS | VSTTS | TVTPVS |
| TXSPVPPS | TXSXPVPS | TTXPVS | TTTPPVS |
| TXSXXPVS | TXTPPPPS | TTPPPVS | TVPXTPS |
| VSXPXTPS | VSXXPXTS | VSSXPVPS | VVPTPPPS |
| VTPPPPPS | VSPVPPPS | TSTPPPSS | TXSXPXPS |
| VSXPVPPS | VSXXPVPS | TSXXPVTS | TXPXSPVS |
| VSXXXPVS | | VXPPPPPS | |

Transfer Stimuli

| Grammatical | | Nongrammatical | |
| --- | --- | --- | --- |
| Similar | Nonsimilar | Similar | Nonsimilar |
| TVPS | TVS | VXS | VVPXS |
| VTPPS | VSPVS | VVPS | TPVPPS |
| VSPVPS | TXSPXTPS | TXTTS | TSPPPS |
| VSXPVS | TXSPXTS | TVPPPVS | VXSPVSS |
| TXSPXTS | VSPXTPPS | TVSPVPPS | VVSPVPS |
| VSPXTPS | VSPXSPVS | TSXXXPVS | TSXSPXTS |

## APPENDIX B
### Experiment 2 Stimuli

Training Stimuli

| Grammar 1 | | Grammar 2 | |
| --- | --- | --- | --- |
| VTS | TXTS | VVS | VVPS |
| VTPS | TVPPS | TPPTS | TTXVS |
| TXTPS | TVPPPS | VXXVS | TPPPTS |
| TXSPVS | TXTPPS | TPTXVS | TTXXVS |
| VSPXTS | VTPPPS | TTXVPS | VXXXVS |
| TVPPPPS | TXSPVPS | TPPPPTS | TPTXXVS |
| TXSXPVS | TXTPPPS | TTXXXVS | TPTXVPS |
| VSXPXTS | VTPPPPS | VXVPXVS | VXXXXVS |
| VSPVPPS | VSXPVPS | VVPXVPS | VXXXVPS |
| VSXXPVS | TVPPPPPS | TPPPPPTS | TPPTXXVS |
| TXSPVPPS | TXSXPVPS | TPTXXXVS | TTXXXXVS |
| TXSXXPVS | TXTPPPPS | TPPTXVPS | TPTXXVPS |
| VSXPXTPS | VSXXPXTS | TTXXXVPS | VXVPXXVS |
| VTPPPPPS | VSPVPPPS | VXXVPXVS | VXXXXXVS |
| VSXPVPPS | VSXXPVPS | VVPXXVPS | VXVPXVPS |
| VSXXXPVS | | VXXXXVPS | |

First Transfer Stimuli (Grammatical)

| Grammar 1 | | Grammar 2 | |
| --- | --- | --- | --- |
| Similar | Nonsimilar | Similar | Nonsimilar |
| TVPS | TVS | TPTS | TTS |
| VTPPS | VSPVS | VXVS | VXVPS |
| VSPVPS | TXSPXTPS | VXXVPS | VVPXVS |
| VSXPVS | TXSPXTS | TPPTXVS | TPPPTXVS |
| TXSPXTS | VSPXTPPS | TTXXVPS | VVPXXXVS |
| VSPXTPS | VSPXSPVS | VVPXXVS | TTXVPXVS |