



# Visual artificial grammar learning by rhesus macaques (*Macaca mulatta*): exploring the role of grammar complexity and sequence length

Lisa A. Heimbauer<sup>1</sup> · Christopher M. Conway<sup>2</sup> · Morten H. Christiansen<sup>3</sup> · Michael J. Beran<sup>2</sup> · Michael J. Owren<sup>4</sup>

Received: 12 July 2017 / Revised: 20 January 2018 / Accepted: 28 January 2018 / Published online: 12 February 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Humans and nonhuman primates can learn about the organization of stimuli in the environment using implicit sequential pattern learning capabilities. However, most previous artificial grammar learning studies with nonhuman primates have involved relatively simple grammars and short input sequences. The goal in the current experiments was to assess the learning capabilities of monkeys on an artificial grammar-learning task that was more complex than most others previously used with nonhumans. Three experiments were conducted using a joystick-based, symmetrical-response serial reaction time task in which two monkeys were exposed to grammar-generated sequences at sequence lengths of four in Experiment 1, six in Experiment 2, and eight in Experiment 3. Over time, the monkeys came to respond faster to the sequences generated from the artificial grammar compared to random versions. In a subsequent generalization phase, subjects generalized their knowledge to novel sequences, responding significantly faster to novel instances of sequences produced using the familiar grammar compared to those constructed using an unfamiliar grammar. These results reveal that rhesus monkeys can learn and generalize the statistical structure inherent in an artificial grammar that is as complex as some used with humans, for sequences up to eight items long. These findings are discussed in relation to whether or not rhesus macaques and other primate species possess implicit sequence learning abilities that are similar to those that humans draw upon to learn natural language grammar.

**Keywords** Artificial grammar learning · Sequence learning · Statistical learning · Rhesus macaques

## Introduction

One of the impressive characteristics of human language acquisition is the degree to which structure is implicitly abstracted through experience with the environment. There is now good reason to believe that at least some of language learning in humans is accomplished by the more general ability to implicitly process statistical input patterns (e.g., Cleeremans et al. 1998; Conway et al. 2010; Kidd 2012;

Kidd and Arciuli 2015; Misyak and Christiansen 2012; Peruchet and Pacton 2006; Reber 1967; Redington and Chater 1997; Saffran 2003; Saffran et al. 1996). Studies with human participants have indicated that this kind of sequence learning is also used for nonlanguage domains, such as the learning of visual or visuo-motor patterns (e.g., Cleeremans and McClelland 1991; Fiser and Aslin 2001, 2002; Hunt and Aslin 2001; Nissen and Bullemer 1987). This general sensitivity to structured patterns, regardless of domain or sensory modality, has also been revealed in other animal species (see Santolin and Saffran 2017).

For instance, it has been demonstrated that pigeons (*Columba livia*) can learn artificial grammars produced using colored letters (Herbranson and Shimp 2008) and that they can abstract these complex regularities to generalize what they learned to novel stimuli strings. Bengalese finches (*Lonchura striata var. domestica*), which use syntactically complex songs to communicate, have the ability to use artificial grammar rules to discriminate songs (Abe and Watanabe 2011). Additionally, after learning the underlying

✉ Lisa A. Heimbauer  
lisa.heimbauer@gmail.com

<sup>1</sup> Division of Liberal Arts and Sciences, State University of New York at Delhi, Evenden Tower Room 715, Delhi, NY 13753, USA

<sup>2</sup> Georgia State University, Atlanta, GA, USA

<sup>3</sup> Cornell University, Ithaca, NY, USA

<sup>4</sup> OSV Acoustical Associates and Emory University, Atlanta, GA, USA

grammatical rules of the synthesized syllable strings, these birds then discriminated novel strings that were produced using the learned rules.

Most recently, Spierings and ten Cate (2016) reported that budgerigars and zebra finches differed in how they generalized what they learned about triplets of song elements. That paper was a first attempt at investigating artificial grammar learning in budgerigars, a parrot species that have an open-ended learning capacity with more vocal variation than zebra finches. Although both species were able to learn the underlying structure of the sound strings, the zebra finches generalized using positional information, while the budgerigars revealed that they could abstract the rules they learned and apply them to novel stimuli.

Nonhuman primates also have demonstrated sensitivity to structured patterns (e.g., Attaheri et al. 2015; Conway and Christiansen 2001; Endress et al. 2010; Hauser et al. 2001; Heimbauer et al. 2012; Locurto et al. 2009, 2010, 2015; Newport et al. 2004; Ravignani et al. 2013; Saffran et al. 2008; Sonnweber et al. 2015; Stobbe et al. 2012; Terrace et al. 2003; Wilson et al. 2013—see Milne et al. 2018, for a review). However, one important and still unanswered question is the extent to which nonhumans demonstrate the capability for learning the underlying structure of complex grammars equivalent to human learning. One way to quantify complexity in artificial grammars is to recognize the distinction between phrase-structure grammars and finite-state grammars (Chomsky 1956, 1959). A phrase-structure grammar (or the equivalent term, context-free grammar) entails center embedding of units that form nested dependencies; this type of rule system is more complex than a finite-state grammar, which consists of local organizational principles only, with statistical regularities limited to neighboring units or connected “states” (for an introduction to the Chomsky hierarchy, see Fitch and Friederici 2012; Fitch et al. 2012; Jäger and Rogers 2012). Although some of the initial evidence appeared to suggest that nonhuman primates were incapable of learning the more computationally demanding, phrase-structure grammar (Fitch and Hauser 2004), more recent research seems to suggest they may have this capability after all (Rey et al. 2012; Stobbe et al. 2012).

Importantly, though, it is unclear whether and how natural language maps onto any level of the Chomsky hierarchy (see Christiansen and Chater 2016, for discussion). Indeed, it has been suggested that human language use may be best characterized by relatively simple computational processes that may be more akin to finite-state systems than their more complex context-free counterparts (e.g., Christiansen and Chater 2015; Frank and Bod 2011; Karlsson 2010; Petersson 2005). From this perspective, it makes sense to focus more closely on a detailed study of finite-state sequence learning abilities in nonhuman primates as a possible precursor for human language learning (see also Petkov and Wilson

2012). Unfortunately, across different studies, different types of artificial grammar learning tasks have been used that vary on a number of different factors (e.g., grammar complexity, stimulus modality, response demands, number of training sessions, number of training exemplars, and sequence length). If different studies use different species using dissimilar learning tasks, and species differences are observed, it is difficult to know the extent to which the differences are due to a discrepancy in the cognitive capabilities of the different species versus differences in the demands of the task itself. What is needed is a way to quantify complexity in artificial grammar learning and then to probe the limits of nonhuman learning by testing their learning for highly complex structures.

Therefore, to better understand the role of grammar complexity (and the related variable sequence length) in artificial grammar learning by nonhuman primates, the goal of the current study was to examine learning by rhesus macaques (*Macaca mulatta*) using a finite-state grammar that was more complex than most others previously used with nonhumans. The grammar task consisted of exposure to statistical-based visual sequences followed by a test of generalization to novel sequences produced using the same statistical patterns. We used Wilson et al.’s (2013) complexity metric to design a grammar that was more complex than nearly all others used in previous nonhuman studies, on a par with those used with humans. Furthermore, because most nonhuman animal studies have used sequences that are much shorter in length than those typically used with humans, a second goal was to vary the length of sequences systematically to examine the effect on learning. In a previous study (Heimbauer et al. 2012), the same animals as in the current study had learned a fixed visual sequence similar in length to those humans are tested with, but the monkeys had not been tested on their ability to learn and generalize grammar-like statistical patterns. Before describing the current study in detail, we first discuss the notion of grammar complexity in artificial grammar learning.

### Grammar complexity in artificial grammar learning

Within the level of finite-state grammars, there can be a wide range of structural complexity, from the use of fixed repeating sequences or chains to more complex and variable probabilistic patterns that are less predictable (Conway and Christiansen 2001). Wilson et al. (2013) developed a metric to quantify the complexity of different types of finite-state grammars (for other complexity metrics, see Pothos 2010; Schiff and Katan 2014). Wilson et al.’s (2013) metric assessed the complexity of an artificial grammar along two dimensions: (1) the number of unique stimulus classes or elements that contribute to the grammar structure and (2) the degree of predictability or determinism in the structure.

The value for the second dimension is calculated by the following equation:

$$LP = (C + 1)/T$$

where LP is linear predictability,  $C$  is the number of unique stimulus elements or structural classes, and  $T$  is the number of unique transitions between stimulus elements. The value for the denominator ( $T$ ) is obtained by creating a list of all of the possible transitions between stimulus elements (Petkov and Wilson, personal communication, May 23, 2016), which also includes transitions between elements within the grammar as well as involving the start and end states. In this formula, a linear predictability of 1.0 represents perfect determinism such as a fixed sequence that is perfectly predictable, whereas lower values represent increasingly higher levels of unpredictability and therefore a higher level of complexity (see Fig. 1 for examples).

To further illustrate these parameters, consider Reber's (1967) classic study. In this grammar, there are five different stimulus elements (V, P, T, X, and S) and 17 unique transitions in the grammar (see Table 1). The linear predictability of this grammar is calculated as  $(5 + 1)/17 = 0.3529$ . In Table 1, we have reproduced Wilson et al.'s (2013) calculations of other human and nonhuman artificial grammar learning studies, updated to include a number of new studies not listed by Wilson et al., and focusing on recent studies with nonhuman primates using finite-state grammars. In addition to the two complexity parameters, we also included additional information including the species tested, the stimulus modality, whether the grammar allows repetitions, and the lengths of the training exemplars. The studies are listed in increasing complexity based on the linear predictability metric. We included all of the finite-state grammar learning studies we know of that have been conducted with nonhuman primates. We also included a sample of human finite-state grammar learning studies for comparison, as well as details for the three experiments reported in the current manuscript.

As depicted in Table 1, out of the seven studies involving finite-state grammar learning by nonhuman primates, six of them (Attaheri et al. 2015; Endress et al. 2010; Heimbauer et al. 2012; Locurto et al. 2013; Saffran et al. 2008; Wilson et al. 2013) incorporated grammars with linear predictability values of 0.5 or higher, which are considerably simpler than most finite-state grammars used in human studies, for instance, by Gomez and Gerken (1999), Conway et al. (2010), Reber (1967), Knowlton and Squire (1996), and Jamieson and Mewhort (2005). Finite-state grammars used in human studies typically have linear predictability values below 0.4, with the more complex ones being lower than 0.3. The only study we know of to use a grammar of such complexity with nonhuman primates was Locurto et al. (2015), who demonstrated learning

of a finite-state grammar with cotton-top tamarins using visual-motor sequences. However, one limitation with this particular grammar is that it allows stimulus repetitions in the generated sequences. Because repetitions are known to be highly salient to learners (Endress et al. 2007), it is likely that this simplified the learning situation. Thus, it is currently unknown to what extent nonhuman primates are able to demonstrate learning of a complex grammar that does not contain stimulus repetitions.

The grammar used in the present study (loosely based on Jamieson and Mewhort 2005) has certain advantages over other artificial grammars that are commonly used. First, unlike most other grammars, including the classic Reber's (1967) grammar, there are no positional constraints. That is, each element of the grammar can occur at any position with equal frequency, eliminating the possibility that specific elements would only occur at the beginning or the end of sequences. Second, there are also no constraints on sequence length. To prevent sequence length from becoming a confound and to allow us to systematically examine learning for sequences of varying lengths, a large set of stimuli can be generated at a particular length. This grammar has the added benefit of being relatively complex in terms of linear predictability (0.3125), on a par with some of the more complex grammars used in human studies and as complex as the Locurto et al.'s (2015) study. However, unlike Locurto et al., the grammar in the current study does not allow stimulus repetitions. This grammar also is able to produce a large set of sequences, much larger than that used in any other nonhuman artificial grammar learning study that we know of. At sequence length of 8, the grammar can produce 512 sequences, creating a relatively formidable learning challenge and making it less likely—though not impossible—that any observed learning effects are due to memorizing individual sequences. In comparison, the studies by Saffran et al. (2008), Wilson et al. (2013), Attaheri et al. (2015), and Endress et al. (2010) used grammars that could only produce a small set of training exemplars (12 or less).

To summarize, the aim of the current study was to assess the learning abilities of rhesus macaques using an artificial grammar that equaled the complexity of most of the grammars used with humans. Furthermore, the previous studies with nonhumans have generally been limited to relatively short (three- to six-item) input sequences. The current experiments were conducted to test nonhuman primates for grammar-like learning and generalization abilities, using a contingency-based reward schedule and relatively long (eight-item) sequences. An additional research question concerned the extent to which these monkeys would display stimulus positional learning effects within the sequences.

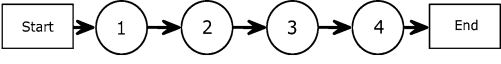
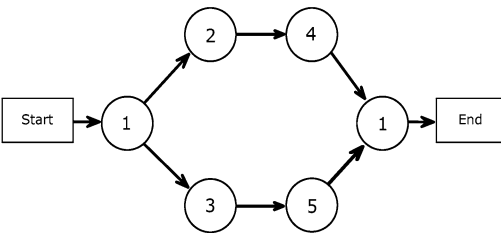
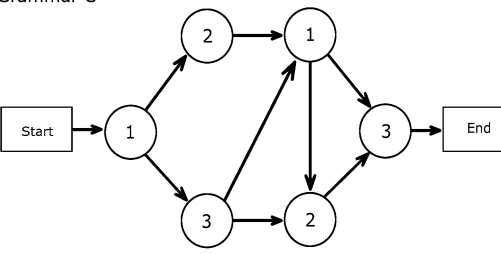
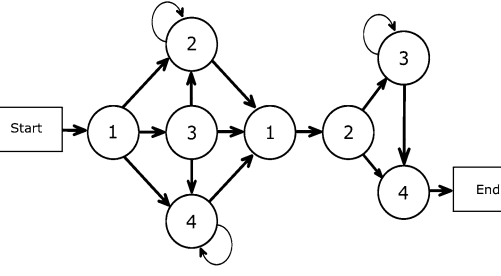
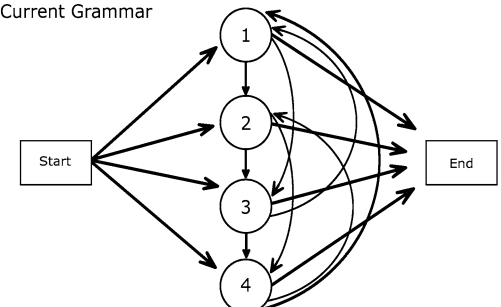
Grammar	Unique Stimulus Classes	Unique Transitions	Linear Probability
<p>Grammar A</p> 	<p>(n=4) 1,2,3,4</p>	<p>(n=5) Start → 1 1 → 2 2 → 3 3 → 4 4 → End</p>	<p><math>LP = \frac{(4+1)}{5} = 1.0</math></p>
<p>Grammar B</p> 	<p>(n=5) 1,2,3,4,5</p>	<p>(n=8) Start → 1 1 → 2 1 → 3 2 → 4 3 → 5 4 → 1 5 → 1 1 → End</p>	<p><math>LP = \frac{(5+1)}{8} = 0.75</math></p>
<p>Grammar C</p> 	<p>(n=3) 1,2,3</p>	<p>(n=8) Start → 1 1 → 2 1 → 3 2 → 1 3 → 1 3 → 2 2 → 3 3 → End</p>	<p><math>LP = \frac{(3+1)}{8} = 0.5</math></p>
<p>Grammar D</p> 	<p>(n=4) A,B,C,D</p>	<p>(n=15) Start → 1    3 → 2 1 → 2    3 → 4 2 → 2    3 → 1 1 → 4    2 → 3 4 → 4    2 → 4 1 → 3    3 → 3 2 → 1    4 → 3 4 → End    4 → 1</p>	<p><math>LP = \frac{(4+1)}{15} = 0.33</math></p>
<p>Current Grammar</p> 	<p>(n=4) 1,2,3,4</p>	<p>(n=16) Start → 1    1 → 3 Start → 2    2 → 4 Start → 3    3 → 1 Start → 4    4 → 2 1 → 2    1 → End 2 → 3    2 → End 3 → 4    3 → End 4 → 1    4 → End</p>	<p><math>LP = \frac{(4+1)}{16} = 0.3125</math></p>

Fig. 1 Examples of artificial grammars that differ in linear predictability values are shown. The grammar at the bottom is the one used in the current study

**Table 1** Summary of parameters used in human and nonhuman artificial grammar learning studies

Study	Species	Modality	# Unique stimulus elements or classes	# Unique transitions	Repetitions	Sequence lengths tested	Linear predictability
Locurto et al. (2013), Exp. 1 and 2	Cotton-top tamarins	Visual (spatial positions)	5	6	N	5 (continuous sequence)	1
Saffran et al. (2008), Exp. 3 (“P-language”)	Cotton-top tamarins	Auditory (non-words)	5	11	N	3–6	0.5454
Wilson et al. (2013)	Rhesus macaques and common marmosets	Auditory (non-words)	5	11	N	3–6	0.5454
Attaheri et al. (2015)	Rhesus macaques	Auditory (non-words)	5	11	N	3–6	0.5454
Saffran et al. (2008) (Exp. 3 (“NP-language”))	Cotton-top tamarins	Auditory (non-words)	5	11	N	3–5	0.5454
Endress et al. (2010), Exp. 1	Chimpanzees	(grunts, screams)	3	8	Y	6	0.5
Heimbauer et al. (2012), Exp. 2	Rhesus macaques	Visual (spatial positions)	4	10	N	5–8 (continuous sequence)	0.5 (for length 8)
Gomez and Gerken (1999)	Humans (infants)	Auditory (non-words)	5	16	Y	3–6	0.375
Conway et al. (2010), Exp. 1	Humans (adults)	Visual (spatial/color locations)	4	14	N	3–8	0.375
Reber (1967), Exp. 1	Humans (adults)	Visual (letter strings)	5	17	Y	6–8	0.3529
Knowlton and Squire (1996), Grammar A	Humans (adults)	Visual (letter strings)	4	15	Y	3–6	0.3333
Locurto et al. (2015)	Cotton-top tamarins and pigeons	Visual (spatial positions)	4	16	Y	Continuous sequence	0.3125
Current Study, Exp. 1	Rhesus macaques	Visual (spatial positions)	4	16	N	4	0.3125
Current Study, Exp. 2	Rhesus macaques	Visual (spatial positions)	4	16	N	6	0.3125
Current Study, Exp. 3	Rhesus macaques	Visual (spatial positions)	4	16	N	8	0.3125
Jamieson and Mewhort (2005), Exp. 1 and 2	Humans (adults)	Visual (color sequences)	6	24	N	8	0.2917
Conway et al. (2010), Exp. 2	Humans (adults)	Auditory (non-words)	4	19	Y	4–8	0.263

We have elected to omit from the table a number of artificial grammar learning studies with nonhumans that have involved structural relationships between just two stimulus classes (e.g., A and B; Fitch and Hauser 2004; Hauser and Glynn 2009; Neiworth 2013; Ravignani et al. 2013; Stobbe et al. 2012; Sonnweber et al. 2015). According to Wilson et al.’s (2013) calculations, these types of grammars are actually quite simple, with only two stimulus classes and linear predictability values that range between 0.6 and 0.8 (see Wilson et al. 2013 for further discussion)

## Current study

In a previous experiment, we reported that rhesus macaques tracked a fixed, visual, eight-item sequence and responded faster at all positions as compared to random sequences using a joystick-based, symmetrical-response serial reaction time (SR-SRT) task (Heimbauer et al. 2012). In this task, a colored circle could appear in one of eight locations

arranged in a circle, equidistant from the starting point to allow for unbiased response measurement. We now tested two of these same monkeys using this SR-SRT paradigm for the ability to learn eight-item sequences generated from a finite-state grammar and to generalize this knowledge to novel sequences produced from the same grammar. Specifically, we used an artificial grammar similar in design to that used by Jamieson and Mewhort (2005), which builds linearly

in such a way that any subsequent item in the sequence is at least partly predictable based on the current item. Each of the four stimulus elements could be followed by one of two other possible elements. In order to test for broad sensitivity to this grammar, a sufficiently large number of sequences were used to ensure that the participants were not simply learning individual sequences by rote, but showing sensitivity to the underlying statistical pattern that all the sequences had in common.

When a monkey was successful at the discrimination task, it was tested for generalization to novel sequences conforming to the same statistical pattern (i.e., generated by the grammar). In the first experiment, we used sequences with a length of four (L4). In the second experiment, we used sequences of L6, and in the third experiment, L8. Increasing the sequence lengths over experiments not only increased the difficulty for tracking each individual sequence, but also exponentially increased the number of sequences that could be presented in each phase of the experiment, thus presenting a formidable learning challenge. L4 represented the shortest length at which testing could occur with this particular grammar, producing 32 possible combinations. L6 created 128 possible sequences, and L8 produced 512 sequences. The goal was therefore to test the monkeys at each length, for evidence of grammar learning and then assessing whether generalization was displayed in each case. Generalization involved comparing performance latencies between the learned sequences (“Familiar”) to new sequences generated from the same artificial grammar used in the learning phase (“Novel-Familiar”) and to sequences generated by a different grammar that a monkey was not previously exposed to (“Unfamiliar”) but that had the same items in it. Table 2 displays the number of test sequences used in each of the three experiments.

In addition, comparison of performance at each position between statistically constrained and random sequences was also examined to ascertain what the monkeys were learning about element positions. It was not expected that any learning would be displayed at Position 1 (apart from motor-related practice effects) because the first circle of a sequence could occur at any location; however, if an animal displayed successful learning of the transitional probabilities of the

grammar, learning at each subsequent position (Positions 2 through 8) should be observed.

## General method

Testing used the SR-SRT task (Heimbauer et al. 2012). In this paradigm, subjects used a joystick to track colored circles that appeared one at a time in fixed locations arranged in a circular pattern on a computer screen.

## Subjects

Subjects were two adult male rhesus monkeys, Obi and Luke, housed at the Language Research Center (LRC) at Georgia State University. Obi was five years old and Luke was nine years old when the experiments began. The monkeys were experienced in using a joystick to respond to stimuli presented on a computer monitor, as well as in using the SR-SRT task. They were neither food- nor water-deprived and had continuous access to a rotating schedule of computerized tasks performed for food reward. Each task was presented for blocks of time ranging from four to six hours. The animals participated in the SR-SRT task approximately three to four sessions per week, completing varying numbers of sequence trials depending on their motivation level and the session duration.

## Apparatus

Subjects were tested individually in their home cages using the LRC Computerized Test System (Richardson et al. 1990; Washburn and Rumbaugh 1992), which included a desktop PC and a 17-in. (43.2 cm) super-VGA monitor positioned approximately 24 cm from the home cage behind a transparent Lexan panel. Each monkey had access to its own testing station and controlled a cursor on the monitor using a vertically mounted joystick protruding into the cage. The joystick was centered on the screen horizontally and located roughly level with the bottom of the monitor. Custom software written in Visual Basic 6.0 (Microsoft Corp., Redmond, WA) was used to control sequence trial presentation and collect data, and task performance was tallied using custom scripts. Statistical analyses were conducted using the online statistical calculators VassarStats (<http://faculty.vassar.edu/lowry/VassarStats.html>) and GraphPad QuickCalcs (<http://www.Graphpad.com/quickcalcs/index.cfm>).

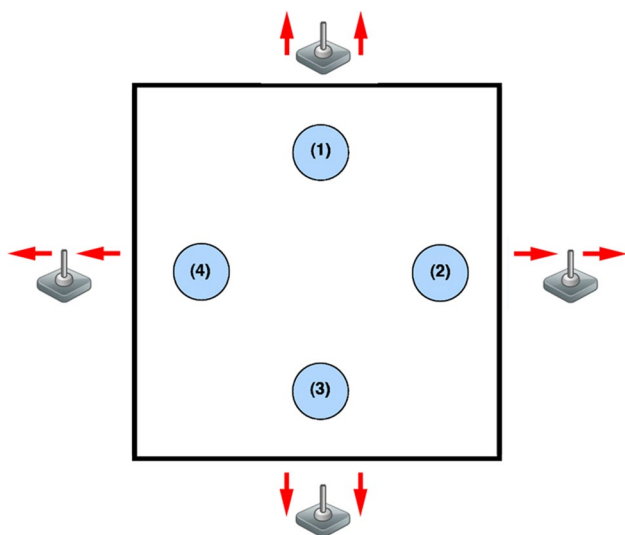
## Procedure

Monkeys performed the SRT task by moving the joystick in the direction of each circle of a sequence as it appeared on the computer monitor. Each circle occupied a

**Table 2** Number of test sequences for Experiments 1 through 3

	Familiar	Novel-Familiar	Unfamiliar
Exp. 1, learning	20		
Exp. 1, generalization	20	4	4
Exp. 2, learning	96		
Exp. 2, generalization	96	16	16
Exp. 3, learning	384		
Exp. 3, generalization	384	64	64





**Fig. 2** The layout of the circle locations on the screen and corresponding joystick movements to track each circle as it appears. Numbers shown in parentheses mark location numbering only and did not appear when circles were displayed

relative “Location” on the screen and a “Position” within the sequence. As shown in Fig. 2, four possible screen locations were used, referred to as 1, 2, 3, and 4. Positions varied from 1 to 8, depending on sequence length. Both the initial learning phase and subsequent generalization phase included statistically constrained (grammar-based) and randomized sequences. The constrained sequences were considered the Familiar sequences, and the randomized sequences were termed Control and used as comparison stimuli for the grammar-based versions. In the generalization phase, two additional sequence types, as described earlier, were presented: Novel-Familiar and Unfamiliar.

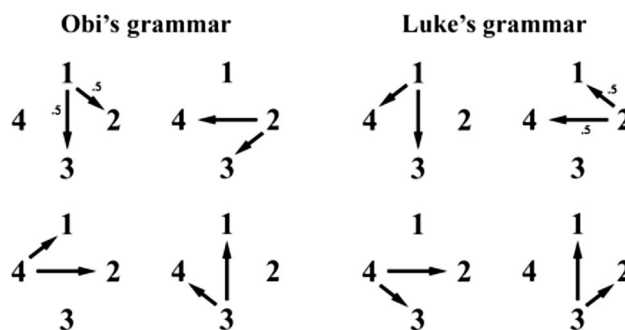
To begin a trial, a monkey used its joystick to move a visible cursor into contact with a start box icon in the middle of the screen. As shown in Fig. 2, a 1.5-in. blue circle appeared in one of four possible locations, arranged in a circular pattern 14.2 cm in diameter. The monkey moved the joystick in the direction of the stimulus, with any detectable deflection registered as a response. Deflection had to fall within 22.5 degrees of the directional heading to the circle, causing it to disappear. Both the cursor and the starting point for the next joystick deflection then reset to the center.

During a sequence, each correct deflection caused a new stimulus to appear in a different location after a 1-s inter-stimulus interval, until the series was complete. Inaccurate deflection or failure to respond to a stimulus within 3 s triggered a 1-s time-out, and the screen went blank. After the time-out, a new sequence began. All types of completed sequences throughout the learning and generalization phases triggered delivery of a 97-mg food pellet, an upswept tone, and a 1-s inter-trial interval. Data collected for each stimulus

presented included sequence type, screen location, position within the sequence, and response latency in ms. Session tallies included sequences completed for each sequence type and median latencies on completed sequences, tabulated by Location and Position.

Only sessions in which a subject initiated at least 100 sequences and completed at least 80% of these sequences were included in analyses. A monkey moved from the learning to the generalization phase in each experiment only when he reached this performance criterion for at least three of four consecutive sessions, while also responding statistically faster to Familiar versus Control sequences in each case. Faster responding was required both overall and for at least two of the individual positions within the sequence, tabulated separately across the sessions for each sequence type. The monkeys were allowed to continue to practice on various sequences for varying lengths of time after criterion performance was met between the learning and generalization phases. The monkeys were also allowed to continue practicing on generalization sessions for 40 to 50 sessions between Experiments 1 and 2 and between Experiments 2 and 3 to allow for adequate learning of the additional sequences in the generalization phases.

The two grammars used are illustrated in Fig. 3 (both have the same underlying structure as the grammar presented at the bottom of Fig. 1). One was assigned to each monkey for the three experiments, and both were based on a given location being followed with equal probability by one of two other designated locations. A location could not follow itself, but could potentially appear in any position within the sequence. Control sequences were pseudo-randomly determined, with the constraint that a location could not follow itself. Learning sessions included 90% grammar-based



**Fig. 3** Schematic illustrations of the grammars used. Each circle location could be followed by either of two others as shown, each with mean occurrence of 50%. Note that if the end state is included in the diagram, then conceivably the transition probabilities are 33%, not 50%. However, for each experiment, length was held constant. This means that for all but the very last item in each sequence, the monkeys were exposed to transition probabilities of 50%, and thus, we believe this is the more accurate value for depicting the statistical probabilities governing transitions within each sequence

sequences and 10% control sequences. Generalization sessions included 70% Familiar, 10% Novel-Familiar, 10% Unfamiliar, and 10% Control sequences. In all sessions, sequences were presented in one of ten different, predetermined, randomized orders. The order used was randomly selected on a session-by-session basis, but the same order was never used in two consecutive sessions.

## Data analysis

Latencies were compared across conditions within each session, both overall and separately by position using one-tailed, unpaired *t* tests, due to the expectation that there would be a directional difference. In the learning phase, comparisons were between Familiar and Control sequences to analyze for overall learning of the Familiar sequences. After overall learning was demonstrated, positional performance was analyzed based on mean positional performance for that session and the following five sessions to look for consistency over subsequent sessions. Generalization sessions were analyzed by conducting repeated-measures ANOVAs and Tukey's HSD for all possible comparisons of the mean RTs for sequences of the four generalization session sequence types: Control, Familiar, Novel-Familiar, and Unfamiliar. The mean RTs for sequence types included Position 1 performance for consistency with our prior work (Heimbauer et al. 2012). To ensure that Position 1 responses were not biasing the overall results, we assessed the RTs at only Position 1 across sequence types to ensure that there were no statistical differences.

Evidence of sensitivity to grammatical statistical structure was deemed to be shown by similar performance on the Familiar and Novel-Familiar sequences, as well as continued faster responding to Familiar grammar-based than randomized Control sequences, Unfamiliar sequences, or both. During the first generalization session, it would be expected that Familiar and Novel-Familiar sequences would show similar reaction times, as would Unfamiliar and Control sequences. Evidence for generalization was looked for in the first generalization session that contained at least 50 completed sequences of each type.

## Experiment 1

The first experiment was designed to investigate whether monkeys responding to visually based sequences generated by a common, underlying statistical-based grammar would show sensitivity (i.e., learning and generalization) to the patterns of L4 sequences. A sequence length of four was used as the shortest testable sequence produced by the simple grammars shown in Fig. 3. The large majority of the resulting sequences were used as Familiar stimuli, thereby

providing the broadest possible exposure to their common structure. While constrained by the relatively small number of sequences available, the rationale was to emphasize learning about the grammar involved rather than the sequences themselves. The consequence was, of course, that only a few were left for testing that learning.

## Procedure

Each of the two grammars used produced 32 possible L4 sequences (see Table 2). Four of these possible sequences were identical for both grammars, namely for Locations 1313, 2424, 3131, and 4242. These were discarded, leaving 28 usable sequences for each grammar. Twenty of these were used as Familiar stimuli and four became Familiar-Novels stimuli. The last four were held back so the experiment could, if necessary, be revised and repeated without extensive retraining. Sequences were assigned to one role or the other on a randomized basis. Each monkey's Novel-Familiar stimuli were used as the other subject's Unfamiliar sequences.

## Results

During learning, two sessions from each monkey did not meet overall performance criteria and were not included in analyses. Over the analyzed sessions, Obi and Luke completed an average of over 750 and 1100 sequence trials, respectively. During session 43 (after a total of 28,956 grammar sequence trials), Obi demonstrated significantly faster, overall responding to Familiar than to Control sequences ( $P < 0.05$ ), in addition to performing significantly faster at individual Positions 2, 3, and 4 of Familiar versus Control sequences during sessions 43 through 48 (Table 3 and Fig. 4a). Luke showed significantly faster overall responding to Familiar versus Control sequences ( $P < 0.01$ ) during session 48 (after a total of 50,205 grammar sequence trials), and was statistically faster at Positions 2, 3, and 4 for Familiar compared to Control sequence positions during sessions 48 through 53 (Table 4 and Fig. 4b). As expected, learning was not displayed at Position 1 because the first circle of a sequence could occur at any location.

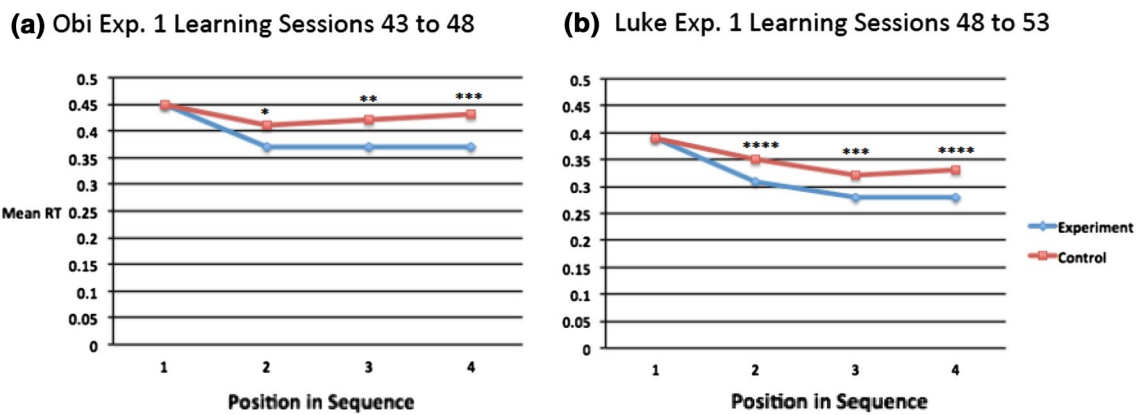
When Novel-Familiar and Unfamiliar sequences were introduced in the generalization phase, ANOVA results for Obi revealed an overall effect of sequence type [ $F(3,177) = 3.31, P < 0.05$ ] with significantly faster reaction times to both Novel-Familiar and Familiar sequences versus Control sequences ( $P < 0.05$ ) during his first session after responding to 60 sequences of each type (see Fig. 5a). Luke's performance was comparable in his first generalization session for 60 sequences of each type, demonstrated by an overall sequence-type effect [ $F(3,177) = 4.25, P < 0.01$ ], with significantly faster performance on both



**Table 3** Obi’s sequence position performance during learning for each experiment

	Exp. 1 (43 through 48)	Exp. 2 (34 through 39)	Exp. 3 (5 through 10)
Position 1	$t(10) = -0.47, P = 0.33$	$t(10) = 1.01, P = 0.17$	$t(10) = 0.59, P = 0.28$
Position 2	$t(10) = 2.36, P < 0.05$	$t(10) = 5.76, P < 0.0001$	$t(10) = 3.0, P < 0.01$
Position 3	$t(10) = 3.56, P < 0.01$	$t(10) = 6.36, P < 0.0001$	$t(6.09) = 2.97, P < 0.01$
Position 4	$t(10) = 4.14, P < 0.001$	$t(10) = 10.13, P < 0.0001$	$t(10) = 4.84, P < 0.001$
Position 5		$t(10) = 7.22, P < 0.0001$	$t(10) = 3.28, P < 0.01$
Position 6		$t(10) = 2.8, P < 0.05$	$t(10) = 2.7, P < 0.01$
Position 7			$t(10) = 3.18, P < 0.01$
Position 8			$t(10) = 2.84, P < 0.01$

Sessions included in the analysis are noted in parentheses



**Fig. 4** Experiment 1 Learning results for Obi (a) and Luke (b); mean RTs for familiar versus control trials at each position (\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; and \*\*\*\* $P < 0.0001$ )

**Table 4** Luke’s sequence position performance during learning for each experiment

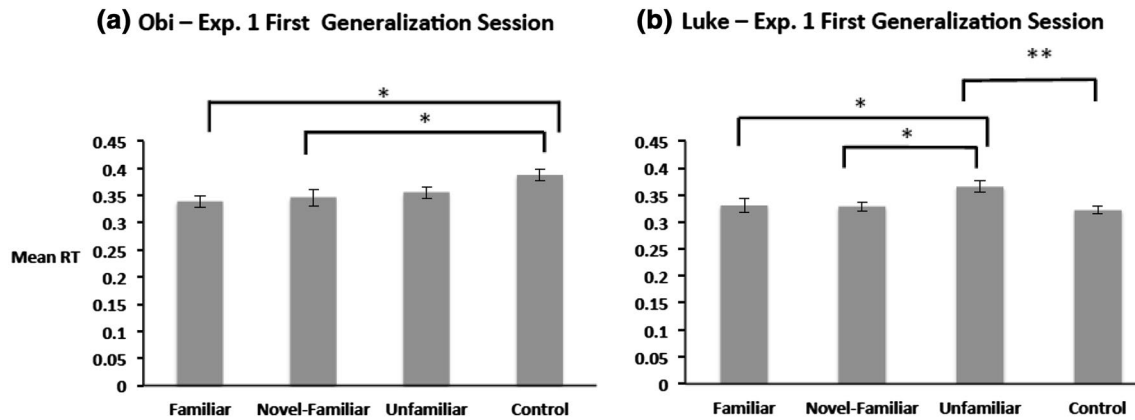
	Exp. 1 (48 through 53)	Exp. 2 (32 through 37)	Exp. 3 (38 through 43)
Position 1	$t(10) = 0.51, P = 0.31$	$t(10) = 0.29, P = 0.39$	$t(10) = 0.05, P = 0.48$
Position 2	$t(10) = 6.55, P < 0.0001$	$t(10) = 0.89, P = 0.20$	$t(10) = 0.14, P = 0.45$
Position 3	$t(10) = 4.46, P < 0.001$	$t(10) = 1.6, P = 0.08$	$t(10) = 1.09, P = 0.15$
Position 4	$t(10) = 10.38, P < 0.0001$	$t(10) = 1.18, P = 0.13$	$t(10) = 1.06, P = 0.16$
Position 5		$t(10) = 1.9, P < 0.05$	$t(10) = 2.15, P < 0.05$
Position 6		$t(10) = 3.42, P < 0.01$	$t(10) = 0.25, P = 0.40$
Position 7			$t(10) = 0.67, P = 0.26$
Position 8			$t(10) = 1.54, P = 0.08$

Sessions included in the analysis are noted in parentheses

Novel-Familiar and Familiar sequences versus Unfamiliar sequences ( $P < 0.05$ ) as shown in Fig. 5b. Luke also demonstrated significantly faster performance on Control versus Unfamiliar sequences ( $P < 0.01$ ). Neither monkey showed a significant difference in performance between the Familiar and Novel-Familiar sequences.

### Discussion

The results showed that both monkeys were sensitive to the underlying statistical patterns of the visual sequences they were exposed to. The learning phase was long for both subjects, with the first evidence of faster overall performance



**Fig. 5** Experiment 1 Generalization (first session) results for Obi (a) and Luke (b); \* $P < 0.05$ ; \*\* $P < 0.01$ . Error bars represent the standard deviation of the means over the analyzed session

in the Familiar versus Control conditions emerging after 43 weeks for Obi and 48 weeks for Luke. The combination of faster performance overall and at individual Positions 2, 3, and 4 on Familiar versus Control sequences, however, demonstrates that the monkeys had formed expectations about sequential relationships among the items. Because the two monkeys were being exposed to differently constructed grammars, it is unlikely that the particular sequences involved were themselves playing a critical role.

Based on the learning phase of the experiment only, the results might reflect learning about the particular 20 sequences being used, rather than their common, underlying structure. The results from the generalization phase argue against this interpretation, providing evidence of immediate comparable performance with novel sequences drawn from each subject's familiar grammar. Although Obi was significantly faster at Novel-Familiar and Familiar sequences when compared to Control sequences, Luke was significantly faster when responding to Novel-Familiar and Familiar sequences as compared to the Unfamiliar sequences. However, neither monkey revealed a difference in performance between the Familiar and Novel-Familiar sequences, supporting the interpretation of generalization of the statistical patterns.

## Experiment 2

Testing with L4 sequences provided evidence that rhesus monkeys can show sensitivity to the statistical structure underlying short visual sequences, including the ability to generalize to novel sequences conforming to the same statistical patterns. However, only 20 different sequences were used during learning, and generalization included only four Novel-Familiar exemplars. The design and outcomes argue against the possibility of rote learning alone accounting for

the findings, but the small number of sequences involved weakens the conclusion. The next step was therefore to test with increased sequence length and a greater total number of possible sequences. Testing at L6 allowed a total of 128 sequences for each, although again with four unusable, overlapping outcomes (i.e., 131313, 242424, 313131, and 424242).

## Procedure

Learning and generalization phases were conducted as in Experiment 1. The same two grammars were used, but now to generate L6 sequences. A total of 96 sequences were used in learning, the four overlapping sequences were discarded, and 12 sequences were held back in case of later need. The remaining 16 sequences were used in generalization testing. Assignment of sequences to condition was again randomized.

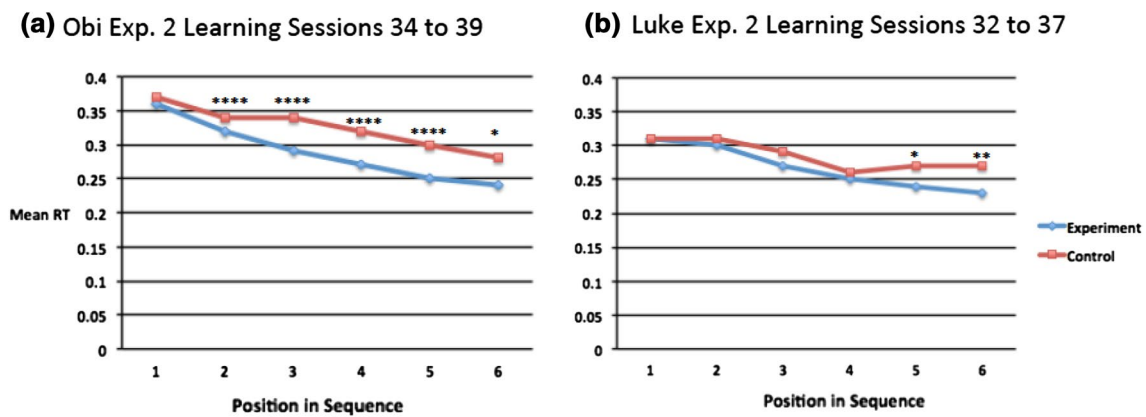
## Results

As mentioned in the “[General method](#)” section, the monkeys were allowed to continue practicing on generalization sessions between experiments to allow for adequate learning of the additional sequences in the generalization phases. To analyze for evidence of sustained learning in these instances, we conducted independent  $t$  tests comparing mean RTs for each of the positions in Familiar sequences in the first generalization session of an experiment to the same position mean RTs in the first session of their learning phase at the next sequence length. When transitioning from Experiment 1 to Experiment 2, Obi responded significantly faster at all positions (1 through 4,  $P < 0.01$ ) in Experiment 2 than he had when meeting learning criteria in Experiment 1. Luke responded the same at Positions 1, 3, and 4 ( $P = ns$ ), and significantly

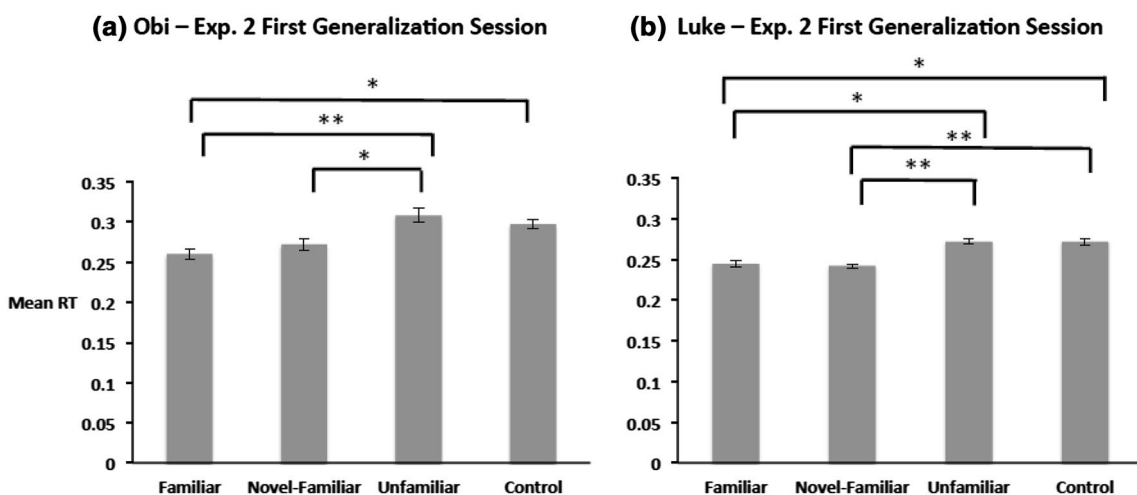
faster at Position 2 ( $P < 0.01$ ) in Familiar sequences in Experiment 2 as when meeting criteria in Experiment 1.

When analyzing learning in Experiment 2, only one session (from Obi) was excluded from analysis due to failure to meet session criteria, and both monkeys completed an average of over 1000 sequences in each session. Both monkeys responded statistically faster to Familiar than to Control sequences: Obi during session 34 (after 29,056 grammar sequence trials), and Luke during session 32 (after 31,057 grammar sequence trials), both at  $P < 0.01$ . At that point in testing, Obi was responding significantly faster at Positions 2 through 6 for Familiar versus Control sequences during sessions 34 through 39 (Table 3 and Fig. 6a), while Luke showed faster responding at Positions 5 and 6 during sessions 32 through 37 (Table 4 and Fig. 6b), with neither monkey revealing a performance difference at Position 1.

Results from generalization testing are shown in Fig. 7a (for Obi) and 7b (for Luke). In this phase, an ANOVA revealed an overall effect of sequence type for Obi after 80 sequences of each type in his first session,  $F(3,237) = 5.31$ ,  $P < 0.01$ . In this session, he was significantly faster responding to both Novel-Familiar and Familiar as compared to Unfamiliar sequences ( $P < 0.05$  and  $P < 0.01$ , respectively), but not as compared to Control sequences. Obi also responded significantly faster responding to Familiar versus Control sequences. Similarly to Obi, Luke demonstrated an overall effect between sequence types in his first session after 80 sequences of each type,  $F(237) = 6.21$ ,  $P < 0.001$ . Similar to his performance in Experiment 1, Luke was significantly faster at Novel-Familiar and Familiar sequences as compared to Unfamiliar sequences ( $P < 0.01$  and  $P < 0.05$ , respectively) and now as compared to Control sequences as well ( $P < 0.01$  and  $P < 0.05$ , respectively). Neither monkey



**Fig. 6** Experiment 2 Learning results for Obi (a) and Luke (b); mean RTs for familiar versus control trials at each position (\* $P < 0.05$ ; \*\* $P < 0.01$ ; and \*\*\*\* $P < 0.0001$ )



**Fig. 7** Experiment 2 Generalization (first session) results for Obi (a) and Luke (b); \* $P < 0.05$  and \*\* $P < 0.01$ . Error bars represent the standard deviation of the means over the analyzed session

demonstrated a significant performance difference between Familiar and Novel-Familiar sequences.

## Discussion

After demonstrating generalization at SL4 in Experiment 1, practicing either sustained or improved training performance for both monkeys when beginning Experiment 2. However, the improvement in RT on Position 1 after additional practice was most likely a basic task-related motor improvement because the first location in a sequence is chosen at random. Similar task-related improvements have also been observed in human SRT tasks (e.g., Misyak and Christiansen 2010).

In this experiment, the results with L6 sequences provided further evidence that the monkeys were sensitive to the grammatical structure of the visual sequences they were tracking. This outcome was not surprising in that the experiment relied on the same grammars the monkeys had already learned. It was nonetheless clear that lengthening the sequences negatively affected both monkeys' performances, with a sizeable number of learning sessions required to bring them back to criterion performance. In addition to showing an overall difference between Familiar and Control sequences, Obi demonstrated faster responding at every position within statistically constrained sequences, with Luke having some difficulty.

As in Experiment 1, Obi and Luke generalized responding to new instances drawn from the grammar they had been trained with in their first generalization session. In this experiment, Luke not only showed generalization by faster responses for statistically constrained Novel-Familiar and Familiar sequences as compared to Unfamiliar sequences, but also to Control sequences as well, which we propose may be evidence of stronger knowledge of the underlying statistical patterns. Additionally, as in Experiment 1, neither monkey showed a difference in response-time performance between the Familiar and Novel-Familiar sequences. Thus, the findings of Experiment 2 provide an even stronger case that the monkeys generalized their knowledge of the grammar-generated statistical patterns, rather than relying on rote memorization of individual sequences.

## Experiment 3

In the third experiment, the sequence length was increased to L8. There were two goals for this manipulation. One was to continue to strengthen the argument that the monkeys had learned about the underlying, grammar-like structure of the sequences they were working with by further increasing the number of exemplars used in both learning and generalization phases. The second goal was to reach a length comparable to those used in testing humans in artificial grammar

experiments (see Cleeremans and McClelland 1991; Conway et al. 2010; Jamieson and Mewhort 2005; Shanks et al. 1997), thereby creating a better basis for comparing performance across species. The same grammars were used again at L8, now allowing 512 possible sequences (with four overlapping and discarded outcomes, 13131313, 24242424, 31313131, and 42424242).

## Procedure

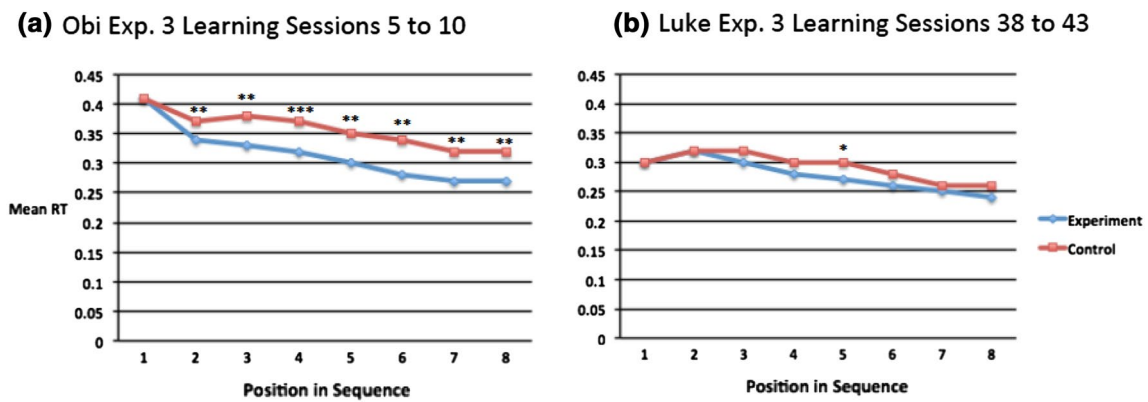
A total of 384 grammar-related sequences were used in the learning phase for each monkey. The four overlapping sequences were discarded from the 128 sequences that remained, of which 64 were selected for generalization testing and 60 were held in reserve.

## Results

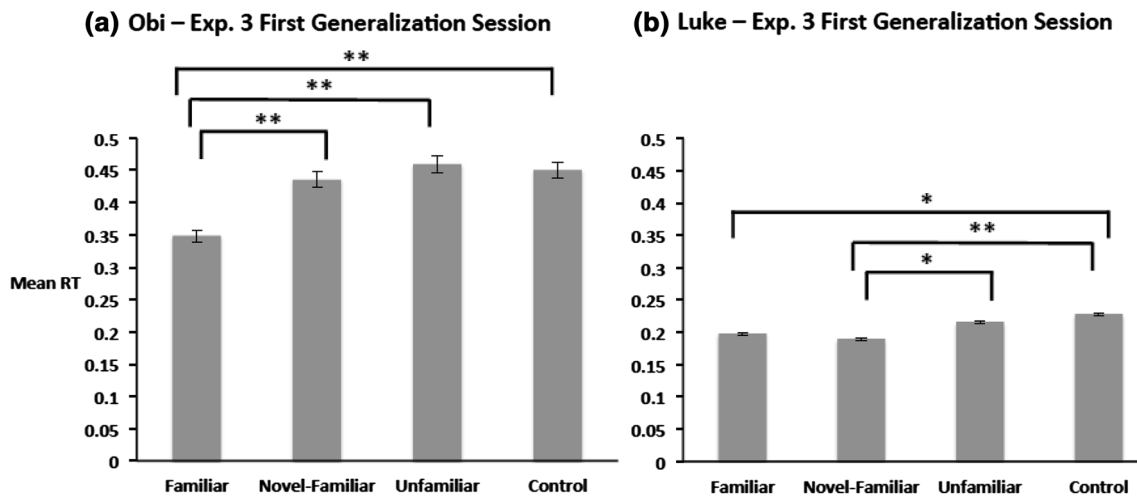
Independent *t* tests comparing mean RTs for each of the positions in Familiar sequences in the first generalization of Experiment 2 to the same position mean RTs in the first session of their learning phase in Experiment 3 were conducted to analyze for any effects of continued practice. The results revealed that Obi performed the same at all positions (1 through 6,  $P = ns$ ) in Familiar sequences in Experiment 3 as when meeting criteria in Experiment 2, and that Luke performed the same at Positions 1 through 5 ( $P = ns$ ) in Experiment 3 as when meeting criteria in Experiment 2. However, at Position 6, Luke was slower ( $P < 0.05$ ) when beginning Experiment 3.

During the SL8 learning phase, two of Obi's L8 sessions were excluded from analysis for not meeting criteria, with Obi completing on average approximately 900 sequences per session and Luke completing approximately 550 per session. Obi responded significantly faster on Familiar versus Control sequences during session 5 (after 3,196 grammar sequence trials), both overall ( $P < 0.01$ ) and at all but Position 1 during sessions 5 through 10 (Table 3 and Fig. 8a). Luke required 32 sessions (after 15,727 grammar sequence trials) to demonstrate learning of Familiar versus Control sequences,  $P < 0.05$ , showing response differences at individual Position 5 between Familiar and Control sequences (Table 4 and Fig. 8b). It should be noted that in this experiment the positional difference was not statistically significant until Sessions 38 through 43 for Luke.

During the first generalization session for which Obi completed at least 50 trials of each sequence type, an ANOVA analyzing performance on the first 60 sequences of each type revealed a significant overall effect,  $F(3,177) = 23.3$ ,  $P < 0.0001$ . Performance was significantly faster on Familiar versus Novel-Familiar, Unknown, and Control sequences, all at  $P < 0.01$  (Fig. 9a). Because Obi did not perform significantly faster on Novel-Familiar sequences versus Unfamiliar



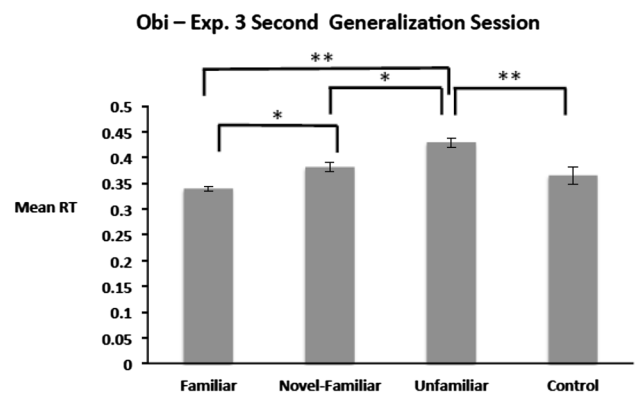
**Fig. 8** Experiment 3 Learning results for Obi (a) and Luke (b); mean RTs for familiar versus control trials at each position (\* $P < 0.05$ ; \*\* $P < 0.01$ ; and \*\*\* $P < 0.001$ )



**Fig. 9** Experiment 3 Generalization (first session) results for Obi (a) and Luke (b); \* $P < 0.05$  and \*\* $P < 0.01$ . Error bars represent the standard deviation of the means over the analyzed session

sequences, we conducted an ANOVA on his second generalization sessions as well. In his second generalization session, the ANOVA results again revealed a significant overall effect,  $F(3,177) = 9.75, P < 0.0001$ . Obi responded significantly faster to Novel-Familiar and Familiar sequences as compared to Unfamiliar sequences,  $P < 0.05$  and  $P < 0.01$ , respectively (Fig. 10), as he did in Experiment 2. Also, in this session his mean response time to Familiar sequences was significantly faster than responses to the Novel-Familiar sequences ( $P < 0.05$ ), and Control sequences were significantly faster than the Unknown sequences ( $P < 0.01$ ).

Luke demonstrated an overall effect of all 55 sequences of each type in his first generalization session,  $F(3,162) = 8.96, P < 0.0001$ . He responded significantly faster to Novel-Familiar and Familiar sequences than to Control sequences,  $P < 0.01$  and  $P < 0.05$ , respectively, in addition to demonstrating statistically faster response times to Novel-Familiar



**Fig. 10** Experiment 3 Generalization (second session) results for Obi; \* $P < 0.05$  and \*\* $P < 0.01$ . Error bars represent the standard deviation of the means over the analyzed session



sequences as compared to Unfamiliar sequences,  $P < 0.05$  (see Fig. 9b).

## Discussion

Again, additional practicing after generalization either increased or sustained learning for both monkeys, with further evidence for possible task-related motor practice effects at Position 1 as discussed earlier. However, it may be that the introduction of two additional positions (7 and 8) slowed Luke down somewhat at Position 6 in his first Experiment 3 session.

As in the previous two experiments, results from the third experiment demonstrated learning and generalization of the artificial grammar-produced sequences. Using sequences with such a long length and an input corpus with such a large number of different exemplars in both learning and generalization phases, makes it highly unlikely that the monkeys were responding based solely on rote memorization. The monkeys' successful performance in the generalization phase, with L8 sequences and a larger number of total sequences, was impressive and equaled or exceeded that found in either of the preceding experiments. Rather than showing any decline with the increased challenge presented in this experiment, generalization performance improved with performance increasing over the three experiments to include faster performance for Novel-Familiar and Familiar compared to both Control and Unfamiliar sequences in some instances. Obi's significantly faster response to the Familiar as compared to the Novel-Familiar sequences appears to reveal the influence of memory for individual instances; he did respond faster to the Novel-Familiar than to the Unfamiliar sequences in the second session, showing that generalization could occur. Overall, we believe that these findings highlight that both monkeys learned the underlying structure of the sequences—perhaps in addition to but not solely based on—remembering specific sequences.

## General discussion

In three experiments, two rhesus macaques used a joystick to respond to the presentation of individual items within sequences of varying lengths (L4, then L6, and lastly L8). Each experiment was designed primarily to test two aspects of learning: the ability to discriminate statistically constrained sequences from random sequences that were not generated from the grammar and the ability to generalize such learning to novel sequences that were generated either from the same grammar or from a new one. All three experiments showed successful learning and generalization by both monkeys, in that their performance on the Familiar grammar-based sequences was significantly faster compared

to the Control or Unfamiliar sequences. Importantly, the monkeys demonstrated these abilities at a sequence length and complexity similar to that used to examine learning in humans (Conway et al. 2010; Jamieson and Mewhort 2005). However, we recognize that faster responding to both the Familiar and Familiar-Novel sequences compared to both the Control and Unfamiliar sequences would indicate the strongest demonstration of learning. This learning effect was only observed in one monkey (Luke) in Experiment 2.

Because the grammar used in the current study was characterized by transitional probabilities, in which any given element could be followed by one of two possible elements with equal probability, successful learning and generalization by the monkeys likely involved sensitivity to such statistical probabilities. Item frequencies in the grammar were balanced, such that all items had an equal probability of occurring. Also, the ordinal position of individual items in each sequence was balanced (any given item could occur at any point in the sequence). Thus, distinguishing between novel sequences that were produced by either the familiar or unfamiliar grammar could not be done by rote memorization of an entire sequence (because all sequences presented in the generalization phase were novel). Neither could discrimination have been accomplished by being sensitive to the frequency of individual items in the grammar (because these were equal across grammars), nor by being sensitive to item position (because all items could occur with equal frequency at any position in the sequence). Distinguishing between the novel sequence types could only be accomplished by learning the predictive nature of the transitional probabilities of the grammatical items experienced during the learning phase and then generalizing those transitional probabilities to the new sequences. On the other hand, some evidence of rote learning of individual items was observed; for instance, Obi demonstrated faster responding to Familiar sequences compared to Familiar-Novel sequences. Previous research has shown that nonhuman primates are able to learn a large number of visual stimuli (e.g., Fagot and Cook 2006) so it is perhaps not surprising that some rote learning of these sequences did occur.

Further evidence for sensitivity to the transitional probabilities inherent in the grammars was demonstrated by the monkeys' performance at individual positions within each sequence. In Experiment 1, both monkeys demonstrated faster response to Positions 2, 3, and 4 in grammar-generated sequences compared to the corresponding positions in the random sequences. Likewise, in Experiments 2 and 3, Obi demonstrated learning of sequential relationships among the items as revealed by his anticipation of elements at all critical positions (Positions 2 through 6 and Positions 2 through 8, respectively). Luke also demonstrated positional learning although not as consistently. In Experiments 2 and 3, Luke did respond faster to Positions 5 and

6 in the grammar-generated sequences than to the random sequences, although this individual difference was not surprising. In our previous experiments with these monkeys, they revealed inconsistencies in positional learning within a fixed sequence (Heimbauer et al. 2012). As was the case with Luke, proximity to reward would be expected to facilitate both learning and motivation toward the end of any given sequence (Locurto et al. 2009, 2010) with observed performance usually becoming faster at the final position. There were also noticeable differences in performance between all Locations for both monkeys throughout the experiments, but there was not an effect for a particular Location. We believe these differences were just a result of the fact that the monkeys were responding to grammars, whether they were familiar, novel-familiar, or unknown.

Regardless of these individual differences in learning, the present results are clear in demonstrating successful learning and generalization of the statistical probabilities contained within visual sequences generated from a relatively complex finite-state grammar. However, this does not necessarily mean that the monkeys learned the grammar as such. Rather, they may have acquired sensitivity to repeated subsequences in the training stimuli, consisting of two or three elements. Indeed, there is a long tradition of using such “chunk” information to explain performance in human sequence learning studies (e.g., Knowlton and Squire 1994—see Perruchet and Pacton 2006, for reviews). That is, sensitivity to statistical regularities need not reflect actual encoding of statistics but may instead result from memorization and generalization over multi-element chunks. Such chunking may stem from well-known constraints on basic memory applied to sequence processing (Christiansen *in press*). Indeed, a recent study of cultural evolution in the laboratory using human participants has shown that language-like structure can emerge through a process of repeated chunking of sequences across “generations” of learners (Cornish et al. 2017). If the monkeys in the current study also engaged in chunking the input sequences, at least in part, then this could potentially explain why some of the random control strings elicited elevated levels of performance because they could in principle have contained one or more familiar chunks by chance, which would have facilitated their processing. It is currently unknown, though, to what extent nonhuman primates employ such chunking processes, and it is important for future studies to compare such potential chunking to human abilities.

It is also worth noting that most of the previous studies with nonhuman primates that have successfully demonstrated sequential learning of a finite-state grammar used much simpler grammars (Attaheri et al. 2015; Endress et al. 2010; Saffran et al. 2008; Wilson et al. 2013; see Santolin and Saffran 2017). The present study and that by Locurto et al. (2015) are the only demonstrations of learning and

generalization of sequences from a more complex artificial grammar. Interestingly, in the present study as well as Locurto et al.’s (2015) visual rather than auditory stimuli were used. It may be purely coincidental that the only demonstrations of learning sequential regularities from a complex grammar involved visual-motor rather than auditory stimuli. On the other hand, it is possible that there are modality constraints affecting sequence processing in nonhuman primates. For example, previous research suggests that nonhuman primates may be more adept at visual temporal and sequential processing compared to auditory processing (Merchant and Honing 2013). Monkeys are quite good at synchronizing motor movements with visual rhythms but have difficulty with auditory-motor synchronization (Zarco et al. 2009). Merchant and Honing (2013) argued that these modality differences are due to the nonhuman primate brain having impoverished auditory-motor connections relative to humans. In terms of human statistical-sequence learning, the sensory modality clearly affects its efficiency and robustness (Conway and Christiansen 2005, 2009; Emberson et al. 2011; see Frost et al. 2015 for a review). However, in the case of humans, auditory sequential learning is generally superior to visual learning, whereas for nonhuman primates, perhaps the opposite is the case (Merchant and Honing 2013). The potential effect of sensory modality on nonhuman primate pattern learning is currently an area in need of further investigation (see Milne et al. 2018 for further discussion).

As generalization of grammatical or statistically based patterns is clearly necessary for developing and understanding human-like language (Kidd 2012), the present study offers the possibility that the necessary cognitive and neural capability to learn, represent, and generalize complex finite-state grammatical structures is phylogenetically old, rather than being human-specific (Wilson et al. 2015a, b). The link between natural language and performance on a visual statistical-sequence learning task such as this one is illustrated by recent evidence, showing that behavioral competence on visual sequence learning tasks is correlated with language competence in humans (Arciuli and Simpson 2012; Conway et al. 2010; Misyak and Christiansen 2010; Misyak et al. 2010). Furthermore, the neural substrates underlying both visual sequential learning of an artificial grammar and visually presented natural language appear to be at least partly coextensive (Christiansen et al. 2012; Petersson et al. 2004; Tabullo et al. 2013). Therefore, it would appear that these two monkeys possess at least some of the appropriate cognitive and neural machinery that may play a role in human-like language.

On the other hand, the grammar used in the current study, while more complex than most others previously used with nonhuman primates, clearly does not have the kind of complexity found in human natural language. Another

consideration to point out is that although both monkeys learned the statistical dependencies of the grammar, it took them a very long time to do so. Even the very complex finite-state grammars used by Jamieson and Mewhort (2005) and Conway et al. (2010) were learnable by human adults within a brief experimental session. The reason for the vast discrepancy in amount of time needed to demonstrate adequate learning is not clear; whether this reveals a true difference in cognitive capacity between the species, or just reflects the particular challenges facing human researchers who work with nonverbal animals, is uncertain (see Conway and Christiansen 2001 for further discussion on the difficulties inherent in making comparisons between humans and nonhumans).

Certainly, the evidence from the present study as well as previous work suggests that humans and some nonhuman primates may share a common learning mechanism for rapidly processing and retaining aspects of serial order and complex structural information (e.g., Endress et al. 2010; Heimbauer et al. 2012; Inoue and Matsuzawa 2009; Locurto et al. 2015; Procyk et al. 2000; Saffran et al. 2008; Terrace et al. 2003; Wilson et al. 2013). However, current evidence also suggests that nonhuman primate communication does not exhibit human-like grammatical complexity. One explanation for this apparent disconnect may be that nonhuman primates are limited relative to humans in their ability to learn more complex structures involving multiple nested nonadjacent dependencies (Conway and Christiansen 2001; de Vries et al. 2011). A recent study testing both humans and macaque monkeys on an auditory artificial grammar containing both adjacent and nonadjacent dependencies appears to support this view (Wilson et al. 2015a, b). If true, this suggests that humans may be drawing on more general processes of the primate cognitive system when passively learning some kinds of statistically based grammar-like patterns, but the more complex processing ability may represent a critical juncture in the evolution of human language acquisition (Christiansen and Chater 2015; Conway and Christiansen 2001).

It may also be possible that statistical-based sequence learning abilities are used differently across different species. Perhaps nonhuman primates use statistical-sequence learning not for communicative purposes, but for other nonlanguage survival functions, such as foraging. Indeed, implicit pattern learning abilities are likely used in an advantageous manner as specific foraging-related studies have demonstrated (de Resende et al. 2008; Lockhard 2008; Macquart et al. 2008; Menzel 2009). What remains an open question is to what extent different species of nonhuman animals (primates and nonprimates) engage in statistical learning and generalization skills for different types of real-world tasks. To further understand the nature and evolution of statistical-sequence learning abilities,

and how they have been co-opted for different types of functions, tasks, and domains, future work would usefully compare nonhuman and human artificial grammar learning performance using a variety of input structures (e.g., adjacent vs. nonadjacent dependencies), sensory modalities (e.g., visual vs. auditory), and task demands (e.g., communicative, navigation, or foraging-related types of tasks).

In sum, these results suggest that the ability that humans draw upon to learn sequence structure may be rudimentary to primates, at least simians. These findings provide the basis for future comparative studies investigating the evolutionary roles of sequence learning in language and other domains.

**Funding** This research was funded in part by National Institutes of Health Grants HD-38051 and HD-060563, an RCALL seed grant, and the College of Arts and Sciences at Georgia State University. Lisa Heimbauer was an RCALL Fellow during completion of this work.

## Compliance with Ethical Standards

**Conflict of interest** Lisa A. Heimbauer declares that she has no conflict of interest. Christopher M. Conway declares that he has no conflict of interest. Morten H. Christiansen declares that he has no conflict of interest. Michael J. Beran declares that he has no conflict of interest. Michael J. Owren is deceased.

**Human and animal rights** All research protocols used in this study with the monkeys were approved by the Georgia State University Animal Care and Use Committee (protocol A09031). All applicable international, national, and/or institutional guidelines for the care and use of animals were followed including those in the Guide for the Care and Use of Laboratory Animals. No human participants were involved in this research.

**Data availability** The datasets analyzed during the current study are available from the corresponding author on reasonable request.

## References

- Abe K, Watanabe D (2011) Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat Neurosci* 14:1067–1173
- Arciuli A, Simpson IC (2012) Statistical learning is related to reading ability in children and adults. *Cogn Sci* 36:286–304
- Attaheri A, Kikuchi Y, Milne AE, Wilson B, Alter K, Petkov CI (2015) EEG potentials associated with artificial grammar learning in the primate brain. *Brain Lang* 148:74–80
- Chomsky N (1956) Three models for the description of language. *IRE Trans Inf Theory* 2:113–124
- Chomsky N (1959) On certain formal properties of grammars. *Inf Control* 2:137–167
- Christiansen MH (in press) Implicit statistical learning: a tale of two literatures. *Top Cogn Sci*
- Christiansen MH, Chater N (2015) The language faculty that wasn't: a usage-based account of natural language recursion. *Front Psychol* 6:1182
- Christiansen MH, Chater N (2016) *Creating language: integrating evolution, acquisition, and processing*. MIT Press, Cambridge

- Christiansen MH, Conway CM, Onnis L (2012) Similar neural correlates for language and sequential learning: evidence from event-related brain potentials. *Lang Cogn Proc* 27:231–256
- Cleeremans A, McClelland JL (1991) Learning the structure of event sequences. *J Exp Psychol Gen* 120:235–253
- Cleeremans A, Destrebecqz A, Boyer M (1998) Implicit learning: news from the front. *Trends Cogn Sci* 2:406–416
- Conway CM, Christiansen MH (2001) Sequential learning in non-human primates. *Trends Cogn Sci* 5:539–546
- Conway CM, Christiansen MH (2005) Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J Exp Psychol Learn* 31:24–39
- Conway CM, Christiansen MH (2009) Seeing and hearing in space and time: effects of modality and presentation rate on implicit statistical learning. *Eur J Cogn Psychol* 21:561–580
- Conway CM, Bauernschmidt A, Huang SS, Pisoni DB (2010) Implicit statistical learning in language processing: word predictability is the key. *Cognition* 114:356–371
- Cornish H, Dale R, Kirby S, Christiansen MH (2017) Sequence memory constraints give rise to language-like structure through iterated learning. *PLoS ONE* 12(1):e0168532
- De Resende BD, Ottoni EB, Fragaszy DM (2008) Ontogeny of manipulative behavior and nut-cracking in young tufted capuchin monkeys (*Cebus apella*): a perception–action perspective. *Dev Sci* 11:828–840
- de Vries MH, Christiansen MH, Petersson KM (2011) Learning recursion: multiple nested and crossed dependencies. *Biolinguistics* 5:10–35
- Emberson LL, Conway CM, Christiansen MH (2011) Timing is everything: changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Q J Exp Psychol* 64:1021–1040
- Endress AD, Dehaene-Lambertz G, Mehler J (2007) Perceptual constraints and learnability of simple grammars. *Cognition* 105:577–614
- Endress AD, Carden S, Versace E, Hauser MD (2010) The apes' edge: positional learning in chimpanzees and humans. *Anim Cogn* 13:483–495
- Fagot J, Cook RG (2006) Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proc Natl Acad Sci* 103:17564–17567
- Fiser J, Aslin RN (2001) Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci* 12:499–504
- Fiser J, Aslin RN (2002) Statistical learning of higher-order temporal structure from visual shape sequences. *J Exp Psychol* 28:458–467
- Fitch WT, Friederici AD (2012) Artificial grammar learning meets formal language theory: an overview. *Philos Trans R Soc B* 367:1933–1955
- Fitch WT, Hauser MD (2004) Computational constraints on syntactic processing in nonhuman primates. *Science* 303:377–380
- Fitch WT, Friederici AD, Hagoort P (2012) Pattern perception and computational complexity: introduction to the special issue. *Philos Trans R Soc B* 367:1925–1932
- Frank SL, Bod R (2011) Insensitivity of the human sentence-processing system to hierarchical structure. *Psychol Sci* 22:829–834
- Frost R, Armstrong BC, Siegelman N, Christiansen MH (2015) Domain generality vs. modality specificity: the paradox of statistical learning. *Trends Cogn Sci* 19:117–125
- Gomez RL, Gerken L (1999) Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70:109–135
- Hauser MD, Glynn D (2009) Can free-ranging rhesus monkeys (*Macaca mulatta*) extract artificially created rules comprised of natural vocalizations? *J Comp Psychol* 123:161
- Hauser MD, Newport EL, Aslin RN (2001) Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* 78:B53–B64
- Heimbauer LA, Conway CM, Christiansen MH, Beran MJ, Owren MJ (2012) A serial reaction time (SRT) task with symmetrical joystick responding for nonhuman primates. *Behav Res Methods* 44:733–741
- Herbranson WT, Shimp CP (2008) Artificial grammar learning in pigeons. *Learn Behav* 36:116–137
- Hunt R, Aslin RN (2001) Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *J Exp Psychol Gen* 130:658–680
- Inoue S, Matsuzawa T (2009) Acquisition and memory of sequence order in young and adult chimpanzees (*Pan troglodytes*). *Anim Cogn* 12:S59–S69
- Jäger G, Rogers J (2012) Formal language theory: refining the Chomsky hierarchy. *Philos Trans R Soc B* 367:1956–1970
- Jamieson RK, Mewhort DJK (2005) The influence of grammatical and local and organizational redundancy on implicit learning: an analysis using information theory. *J Exp Psychol Learn* 31:9–23
- Karlssohn F (2010) Syntactic recursion and iteration. In: van der Hulst H (ed) *Recursion and human language*. Mouton de Gruyter, Berlin, pp 43–67
- Kidd E (2012) Implicit statistical learning is directly associated with the acquisition of syntax. *Dev Psychol* 48:171–184
- Kidd E, Arciuli J (2015) Individual differences in statistical learning predict children's comprehension of syntax. *Child Dev* 87:184–193
- Knowlton BJ, Squire LR (1994) The information acquired during artificial grammar learning. *J Exp Psychol Learn* 20:79–91
- Knowlton BJ, Squire LR (1996) Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *J Exp Psychol Learn* 22:169
- Lockhard E (2008) Effects of foraging sequence on the ability of lambs to consume endophyte-infected tall fescue (alkaloids), birdsfoot trefoil (tannins), and alfalfa (saponins). Unpublished doctoral dissertation, Utah State University, Utah
- Locurto C, Gagne M, Levesque K (2009) Implicit chaining in cotton-top tamarins (*Saguinus oedipus*). *J Exp Psychol Behav Process* 35:116–122
- Locurto C, Gagne M, Nutile K (2010) Characteristics of implicit chaining in cotton-top tamarins (*Saguinus oedipus*). *Anim Cogn* 13:617–629
- Locurto C, Dillon L, Collins M, Conway M, Cunningham K (2013) Implicit chaining in cotton-top tamarins (*Saguinus oedipus*) (with elements equated for probability of reinforcement). *Anim Cogn* 16:611–625
- Locurto C, Fox M, Mazella A (2015) Implicit learning in cotton-top tamarins (*Saguinus oedipus*) and pigeons (*Columba livia*). *Learn Behav* 43:129–142
- Macquart D, Latil G, Beugnon G (2008) Sensorimotor sequence learning in the ant *Gigantiops destructor*. *Anim Behav* 75:1693–1701
- Menzel R (2009) Serial position learning in honeybees. *PLoS ONE* 4:e4694
- Merchant H, Honing H (2013) Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front Neurosci* 7:274
- Milne AE, Wilson B, Christiansen MH (2018) Structured sequence learning across sensory modalities in humans and nonhuman primates. *Curr Opin Behav Sci* 21:39–48
- Misyak JB, Christiansen MH (2010) When 'more' in statistical learning means 'less' in language: individual differences in predictive processing of adjacent dependencies. In: Catrambone R, Ohlsson S (eds) *Proceedings of the 32nd annual cognitive science society conference*. Cognitive Science Society, Austin, pp 2686–2691

- Misyak JB, Christiansen MH (2012) Statistical learning and language: an individual differences study. *Lang Learn* 62:302–331
- Misyak J, Christiansen MH, Tomblin JB (2010) On-line individual differences in statistical learning predict language processing. *Front Psychol* 1:31
- Neiworth JJ (2013) Chasing sounds. *Behav Process* 93:111–115
- Newport EL, Hauser MD, Spaepen G, Aslin RN (2004) Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cogn Psychol* 49:85–117
- Nissen MJ, Bullemer P (1987) Attentional requirements of learning: evidence from performance measures. *Cogn Psychol* 19:1–32
- Perruchet P, Pacton S (2006) Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn Sci* 10:233–238
- Petersson KM (2005) On the relevance of the neurobiological analogue of the finite state architecture. *Neurocomputing* 65–66:825–832
- Petersson KM, Forkstam C, Ingvar M (2004) Artificial syntactic violations activate Broca's region. *Cogn Sci* 28:383–407
- Petkov CI, Wilson B (2012) On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philos Trans R Soc B* 367:2077–2088
- Pothos EM (2010) An entropy model for artificial grammar learning. *Front Psychol* 1:1–13
- Procyk E, Dominey PF, Amiez C, Joseph JP (2000) The effects of sequence structure and reward schedule on serial reaction time learning in the monkey. *Cogn Brain Res* 9:239–248
- Ravignani A, Sonnweber R-S, Stobbe N, Fitch WT (2013) Action at a distance: dependency sensitivity in a New World primate. *Bio Lett* 9:20130852. <https://doi.org/10.1098/rsbl.2013.0852>
- Reber AS (1967) Implicit learning of artificial grammars. *J Verbal Learning Verbal Behav* 77:317–327
- Redington M, Chater N (1997) Probabilistic and distributional approaches to language acquisition. *Trends Cogn Sci* 1:273–281
- Rey A, Perruchet P, Fagot J (2012) Centre-embedded structures are a by-product of associative learning and working memory constraints: evidence from baboons (*Papio Papio*). *Cogn* 123:180–184
- Richardson WK, Washburn DA, Hopkins WD, Savage-Rumbaugh ES, Rumbaugh DM (1990) The NASA/LRC computerized test system. *Behav Res Methods* 22:127–131
- Saffran JR (2003) Statistical language learning: mechanisms and constraints. *Curr Dir Psychol Sci* 12:110–114
- Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–1928
- Saffran J, Hauser M, Seibel R, Kapfhamer J, Tsao F, Cushman F (2008) Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* 107:479–500
- Santolin C, Saffran JR (2017) Constraints on statistical learning across species. *Trends Cogn Sci*. <https://doi.org/10.1016/j.tics.2017>
- Schiff R, Katan P (2014) Does complexity matter? Meta-analysis of learner performance in artificial grammar tasks. *Front Psychol* 5:1084
- Shanks DR, Johnstone T, Staggs L (1997) Abstraction processes in artificial grammar learning. *Q J Exp Psychol* 50A:216–252
- Sonnweber R, Ravignani A, Fitch WT (2015) Non-adjacent visual dependency learning in chimpanzees. *Anim Cogn* 18:733–745
- Spierings M, ten Cate C (2016) Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proc Natl Acad Sci* 113:E3977–E3984
- Stobbe N, Westphal-Fitch G, Aust U, Fitch WT (2012) Visual artificial grammar learning: comparative research on humans, kea (*Nestor notabilis*) and pigeons (*Columba livia*). *Philos Trans R Soc B* 367:1995–2006
- Tabullo A, Sevilla Y, Segura E, Zanutto S, Wainelboim A (2013) An ERP study of structural anomalies in native and semantic free artificial grammar: evidence for shared processing mechanisms. *Brain Res* 1527:149–160
- Terrace HS, Son LK, Brannon EM (2003) Serial expertise of rhesus macaques. *Psychol Sci* 14:66–73
- Washburn DA, Rumbaugh DM (1992) Testing primates with joystick-based automated apparatus: lessons from the Language Research Center's computerized test system. *Behav Res Methods* 24:157–164
- Wilson B, Slater H, Kikuchi Y, Milne AE, Marslen-Wilson WD, Smith K, Petkov CI (2013) Auditory artificial grammar learning in macaque and marmoset monkeys. *J Neurosci* 33:18825–18835
- Wilson B, Kikuchi Y, Sun L, Hunter D, Dick F, Smith K, Thiele A, Griffiths T, Marslen-Wilson WD, Petkov C (2015a) Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nat Commun* 6:8901
- Wilson B, Smith K, Petkov CI (2015b) Mixed-complexity artificial grammar learning in humans and macaque monkeys: evaluating learning strategies. *Eur J Neurosci* 41:568–578
- Zarco W, Merchant H, Prado L, Mendez JC (2009) Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *J Neurophysiol* 102:3191–3202