Behavioral/Cognitive

Auditory Artificial Grammar Learning in Macaque and Marmoset Monkeys

Benjamin Wilson,[1,2] **Heather Slater,**[1,2] **Yukiko Kikuchi,**[1,2] **Alice E. Milne,**[1,2] **William D. Marslen-Wilson,**[3] **Kenny Smith,**[4] and **Christopher I. Petkov**[1,2]

[1]Institute of Neuroscience, and [2]Centre for Behaviour and Evolution, Newcastle University, Newcastle upon Tyne, NE2 4HH, United Kingdom, [3]Department of Psychology, University of Cambridge, Cambridge, CB2 3EB, United Kingdom, and [4]School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, EH8 9AD, United Kingdom

Artificial grammars (AG) are designed to emulate aspects of the structure of language, and AG learning (AGL) paradigms can be used to study the extent of nonhuman animals' structure-learning capabilities. However, different AG structures have been used with nonhuman animals and are difficult to compare across studies and species. We developed a simple quantitative parameter space, which we used to summarize previous nonhuman animal AGL results. This was used to highlight an under-studied AG with a forward-branching structure, designed to model certain aspects of the nondeterministic nature of word transitions in natural language and animal song. We tested whether two monkey species could learn aspects of this auditory AG. After habituating the monkeys to the AG, analysis of video recordings showed that common marmosets (New World monkeys) differentiated between well formed, correct testing sequences and those violating the AG structure based primarily on simple learning strategies. By comparison, Rhesus macaques (Old World monkeys) showed evidence for deeper levels of AGL. A novel eye-tracking approach confirmed this result in the macaques and demonstrated evidence for more complex AGL. This study provides evidence for a previously unknown level of AGL complexity in Old World monkeys that seems less evident in New World monkeys, which are more distant evolutionary relatives to humans. The findings allow for the development of both marmosets and macaques as neurobiological model systems to study different aspects of AGL at the neuronal level.

Introduction

Language is a uniquely human trait with poorly understood evolutionary origins (Bickerton and Szathmary, 2009; Hurford, 2012). Because of its complexity in meaning ("semantics") and structure ("syntax"), natural language cannot be directly investigated in nonhuman animals. However, theoretical work has identified distinct computations related to language that can be comparatively studied (Hauser et al., 2002; Bickerton and Szathmary, 2009; Hurford, 2012). Initial approaches studied referential communication in animals, which has inspired work on how neurons process communication signals (Seyfarth et al., 1980;

Tian et al., 2001). Recently, songbirds have been viewed as promising neurobiological model systems because, like humans and a few other animal species, they are vocal learners and can produce songs with "syntax-like" structure (Berwick et al., 2011). Yet, vocal production learning appears to have occurred by convergent evolution rather than by common descent, since nonhuman primates and most other species have more limited vocal production capabilities (Petkov and Jarvis, 2012). This has raised questions regarding whether nonhuman primates might be able to learn structural patterns with sufficient levels of complexity to provide novel insights on language precursors and their study in animal models.

Artificial grammars (AG) can be created to emulate certain aspects of the structure of natural language or simpler "rule-based" structures that some animals might be able to learn. These can be comparatively studied using AG learning (AGL) paradigms (Fitch and Hauser, 2004). In such studies human participants or nonhuman animals have no *a priori* knowledge about the structure of the AG. Yet, by being habituated to or trained with exemplary sequences of sensory stimuli generated by the AG, the relationship between the elements in the sequence can be acquired [sometimes also referred to as "statistical learning" (Saffran et al., 1996, 1999)]. Differential responses to novel well formed (correct) sequences compared with those that violate the AG structure suggest that some aspect of the AG structure was learned. Although several nonhuman animal AGL studies have been conducted, cross-species comparisons between different nonhuman primates species are needed (Fitch and Hauser, 2004;

Saffran et al., 2008; Petkov and Jarvis, 2012). We asked whether New World monkeys (sharing a last common ancestor, LCA, with humans ~40 million years ago) would have better, comparable, or worse AGL capabilities than Old World monkeys (LCA ~25 million years ago)?

This study first compared different AG structures and animal AGL results within a quantitative parameter space, which identified gaps in our understanding. To address these gaps, we studied New and Old World monkeys (respectively, common marmosets and Rhesus macaques) using a refined AGL approach based on rating videotaped animal responses. We obtained evidence that both species notice violations of the AG, but while macaques show sensitivity to more complex aspects of the AG, the marmosets' responses are based largely on simpler strategies. We developed a novel eye-tracking technique to further investigate the extent of the AGL in individual macaques, the results of which supported the video-coding results and further ruled out simple learning strategies in the macaques.

Materials and Methods

This research study abides by the recommendations of the Weatherall report on "The use of nonhuman primates in research." The study has been approved by the U.K. Home Office and abides by the Animal Scientific Procedures Act (1986) of the United Kingdom.
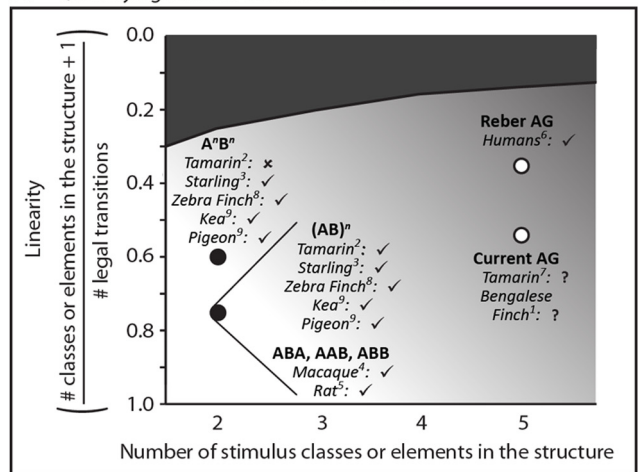
A quantitative parameter space to compare artificial grammar complexity

It is important to quantify some of the dimensions within which AGs can vary, so that different AG structures can be compared or meaningfully varied, rather than being arbitrarily designed or redesigned. The Formal Language Hierarchy and its more recent variants have suggested categorical distinctions between grammars of different levels of complexity (Chomsky, 1957; Berwick et al., 2011). However, a number of groups have emphasized the need for alternative complexity measures to evaluate syntactic complexity (de Vries et al., 2011; Hurford, 2012; Jäger and Rogers, 2012; Petkov and Wilson, 2012). Even "finite-state grammars" (FSGs), holding the lowest place in the Formal Language Hierarchy, can have considerable variability in structural complexity, which we aim to better understand within a quantitative parameter space.

One important variation in complexity between AGs is in the number of stimulus classes or elements that contribute to the AG structure. Although human studies have used a variety of AGs (Reber, 1967; Saffran et al., 2008; Uddén et al., 2012), many studies with nonhuman animals have focused on structural relationships between two stimulus classes: i.e., A and B (Fitch and Hauser, 2004; Friederici et al., 2006; Gentner et al., 2006; Murphy et al., 2008; Hauser and Glynn, 2009). Such AGs require the participants to learn how several stimuli are subdivided into the two classes—e.g., based on salient acoustic features such as the gender of the speaker (Fitch and Hauser, 2004)—before the participants can learn to recognize well formed sequences from the AG structure. These structures are represented by filled circles in the left part of the parameter space in Figure 1A. Other AG studies do not rely on such binary categorization, and instead employ multiple elements that contribute to the structure, which we term "structural elements"; open circles in Figure 1A (Reber, 1967; Saffran et al., 2008; Abe and Watanabe, 2011). Several structural elements typically contribute to the structure of such AGs (Reber, 1967). This can be used to generate a wide variety of sequences without requiring participants to perceptually categorize stimuli into two different classes. Accordingly, the first dimension in Figure 1A is the number of stimulus classes (in reference to studies using AB-type structures) or structural elements (in reference to studies which do not require categorization of stimuli) that contribute to the AG structure.

A second key source of variation between AGs is the degree of predictability or determinism of the structure, reflecting the extent to which each stimulus class or structural element can be predicted by the preceding element(s). The sequence of words or phrases in human language is generally nondeterministic, making it important to understand how far
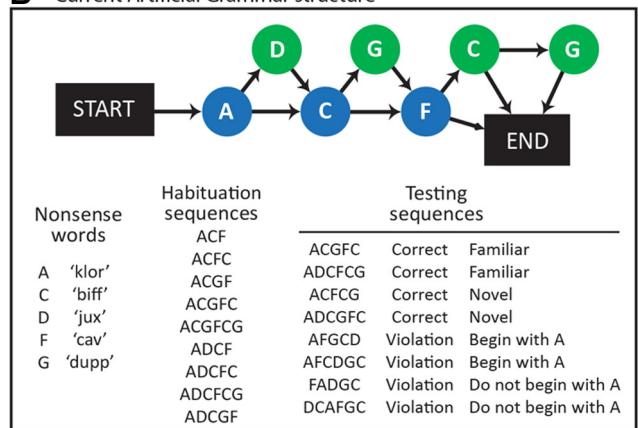

Figure 1. Comparing different AG structures and the current paradigm. **A**, Mapping of AGs previously used to test nonhuman animals, including the original AG designed by Reber (1967). These are plotted as the number of unique stimulus classes (filled circles) or structural elements (open circles) that contribute to the structure as a function of the linearity of the structure (see Materials and Methods). The black line subdividing the shaded regions denotes the maximum possible structural nonlinearity (i.e., random patterns devoid of structure). The checkmarks highlight regions of the parameter space for which there is evidence that the different animal species (labeled text in **A**) can learn that particular level of structural complexity. Crosses or question marks highlight uncertainty regarding whether the labeled species can learn those aspects, see text. Figure references: 1: Abe and Watanabe (2011), 2: Fitch and Hauser (2004), 3: Gentner et al., (2006), 4: Hauser and Glynn (2009), 5: Murphy et al. (2008), 6: Reber (1967), 7: Saffran et al. (2008), 8: van Heijningen et al. (2009), 9: Stobbe et al. (2012). **B**, The AG structure used here contains five unique elements and multiple forward branching relationships. Correct sequences (strings of nonsense words) are generated by following any path of arrows from START to END. Violation sequences do not follow the arrows. The AG was used to create 9 habituation sequences. All experiments began with a habituation phase following by a testing phase. The testing sequences that follow the AG ("Correct") or do not follow the AG ("Violation") are also shown.

nonhuman animals are sensitive to similar properties in the sequences generated by a given AG. The songs of some songbird species, for example, can range from stereotyped and deterministic to much more variable. This can be quantified by calculating their structural linearity (Honda and Okanoya, 1999), given by the following:

Linearity

$$= \frac{\text{Number of stimulus classes or structural elements} + 1}{\text{Number of legal transitions}}.$$

A linearity index of 1.0 describes an entirely predictable, deterministic AG, where each structural element can be preceded and followed by only

one legal transition. The equation above includes transitions between structural elements and also to and from the start or end of the sequence. The number of structural elements considered in this equation contains an additional token so that a manifestly linear AG, e.g., Start $\rightarrow$ A $\rightarrow$ B $\rightarrow$ End, has 2 structural elements (A and B) and three transitions ($\rightarrow$), where Linearity = $(2 + 1)/3 = 1.0$. Thus, the second dimension in Figure 1A is a measure of structural linearity.

Quantifying AG complexity and evaluating animal AGL capabilities

On this two-dimensional space, we first mapped the AG structures containing just two stimulus classes: $(AB)^n$ and A^nB^n. The $(AB)^n$ structures produce sequences of the form ABAB (where $n = 2$) and the A^nB^n structures produce the sequence AABB (Fitch and Hauser, 2004; Gentner et al., 2006). We also mapped three-element long structures based on the A/B classes of stimuli, producing sequences such as ABA, AAB, ABB (Murphy et al., 2008; Hauser and Glynn, 2009). See lower-left area of Figure 1A. The $(AB)^n$ and the A/B structures are relatively linear, with only one transition that is not entirely predictable based on the prior stimulus class, e.g. $(AB)^n$ must begin with A, A is then always followed by B, and B can be followed by either A or "End". Every species tested appears able to learn this type of AG structure, either implicitly or explicitly: songbirds (Gentner et al., 2006; van Heijningen et al., 2009; Stobbe et al., 2012), rodents (Murphy et al., 2008), New and Old world monkeys (Fitch and Hauser, 2004; Hauser and Glynn, 2009). This suggests that many species are capable of learning AG structures based on these relatively linear, predictable, structural relationships.

By comparison, A^nB^n structures (e.g., AABB, where $n = 2$) are more nonlinear since A may be followed by either A or B (Fig. 1A). After training, a number of avian species were able to detect violations in both $(AB)^n$ and A^nB^n structures (up to $n = 4$) (Gentner et al., 2006; van Heijningen et al., 2009; Stobbe et al., 2012). However, tamarin monkeys (a New World monkey species) showed dishabituation responses only to violations of the $(AB)^n$ structure but not to the A^nB^n structure (where $n = 2$) (Fitch and Hauser, 2004). It is unclear whether these differences between monkeys and birds result from the difference between learning by training or habituation (i.e., explicit vs implicit forms of learning) or reflect a genuine cross-species difference in AGL capabilities. However, the results do not provide evidence that tamarin monkeys are able to learn less linear AG structures of this type.

The other AG structures, mapped in the right half of Figure 1A, consist of several structural elements, offering considerable variation in the sequences that can be generated by each AG. For example, the two-stimulus class $(AB)^n$ structure generates the fixed sequence of the form ABAB (where $n = 2$), with the bulk of the learning effort used to identify how the different stimuli fit into the two classes. By comparison, the AG structure in "Reber-like" AGs can produce a variety of sequences and sequence lengths, e.g., "TPTXVS" or "VXVPXXVS" (Reber, 1967). Since several structural elements contribute to the AG, Reber-like structures are typically less linear and deterministic than those that can be generated by two-stimulus class AGs (Fig. 1A). While AGs such as those inhabiting the upper right quadrant in Figure 1A are learned with relative ease by human participants (Reber, 1967; Friederici et al., 2002; Petersson et al., 2012), they have not been tested with nonhuman animals and might prove very difficult for them to learn. For this study of implicit AGL in nonhuman primates we focused on a Reber-like AG developed by Saffran et al. (2008). In terms of linearity, this AG structure (Fig. 1B) falls between the structure used by Reber (1967) and the two-stimulus class structures. This AG can generate sequences of variable length and the order of the elements varies between sequences. The structure contains both optional and obligatory elements including a considerable variety of transitional probabilities between elements (Fig. 1B; Table 1).

Two previous studies have attempted to determine whether nonhuman animals can learn AGs with similarly nondeterministic structure. In the first study, after tamarin monkeys were habituated to sequences generated by the AG, the only evidence for significant dishabituation responses to violations of the AG structure was obtained when the animals were tested with the same "correct" sequences to which they had been habituated (Saffran et al., 2008). Thus, the dishabituation responses of

Table 1. Transitional probabilities between elements and for test sequences

Transition	Transitional probability (TP)	Test sequences	Average TP
Start–A	1.00		
A–C	0.56	ACGFC	0.57[a]
A–D	0.44	ADCFCG	0.59[a]
D–C	1.00	ACFCG	0.54[a]
C–F	0.36	ADCGFC	0.62[a]
C–G	0.43		
C–End	0.21		
F–C	0.56	AFGCD	0.17[b]
F–End	0.44	AFCDGC	0.25[b]
G–F	0.67	FADGC	0.11[b]
G–End	0.33	DCAFGC	0.17[b]

The transitional probability (TP) of every legal transition between elements was calculated based on the frequency of their occurrence within the habituation sequences. Higher TPs represent more common transitions. The average TP of each test sequence is also shown, highlighting the higher average TPs in the [a]correct than in the [b]violation sequences.

these New World monkeys may be based primarily on the novelty of the violation sequences. In our experimental design we incorporated both "familiar" and "novel" correct (well formed) testing sequences to determine whether macaques (Old World monkeys) and/or marmosets (New World monkeys) would distinguish between sequences only on the basis of familiarity. Second, in a study testing Bengalese finches on a related AG structure (Abe and Watanabe, 2011), it has been noted that the testing sequences used differed significantly in their acoustic properties between conditions. All correct sequences were acoustically very similar to each other but the violation sequences differed considerably (Beckers et al., 2012). Thus, the animals could have responded differently to the test sequences based solely on acoustical differences. To address this, our experimental design involved selecting violation sequences that violate the AG structure at multiple positions in the sequences, and we controlled for acoustic differences between correct and violation sequences (see Stimuli, below). Last, to better clarify what parts of the sequence the animals monitor for violations (van Heijningen et al., 2009), this study incorporated two different types of violation sequences: those that "begin with A" (like the well formed, correct sequences) and those that "do not begin with A" (violate the sequence structure from the very first element, Fig. 1B).

Video-coding experiments

Stimuli. Each of the stimulus sequences shown in Figure 1B was created by digitally combining recordings of naturally spoken nonsense words produced by a female speaker based on an AG structure developed by Saffran et al. (2008) (Fig. 1B). The nonsense words were recorded with an Edirol R-09HR (Roland) sound recorder. The amplitude of the recorded sounds was root-mean-square (RMS) balanced. The nonsense word stimuli were combined into habituation and testing sequences using customized Matlab scripts [100 ms interstimulus intervals (ISI)]. The sounds were presented to the animals using Cortex software (Salk Institute) at ~75 dB SPL (calibrated with an XL2 sound level meter, NTI Audio). We confirmed that the power spectrum density of the nonsense word stimuli was well within the audible range of both macaques and marmosets [i.e., at least 30 dB above both species' hearing threshold in the range of 100–5000 Hz (Pfingst et al., 1975, 1978; Lonsbury-Martin and Martin, 1981; Bennett et al., 1983) i.e., at least 30 dB above both species' hearing threshold in the range of ~100–5000 Hz (Seiden, 1958)]. The duration of the naturally spoken nonsense word stimuli within the sequences varied (Klor = 0.64 s; Jux = 0.62 s; Cav = 0.56 s; Biff = 0.40 s; Dupp = 0.39 s). Thus, we confirmed that the duration of the sequences could not be used as a cue, as follows. The correct and violation sequence sets were balanced in the number of elements in the sequences (Fig. 1B) and the mean sequence length ($\pm$SD) of the sequences were comparable: correct sequences, 3.14 (0.42) s; violation sequences, 3.25 (0.28) s. Also, we confirmed that there was no significant difference in sequence sound duration between correct and violation sequences (independent samples t test, $t_{(6)} = 0.435$, $p = 0.68$), or in the

duration of the individual elements present in correct versus violation sequences ($t_{(42)} = 0.609$, $p = 0.55$).

Moreover, further steps were taken in designing the sequence sets to balance for acoustical differences, by either balancing for the presence of the different elements (A, C, D, F, G), as far as possible, or analytically confirming that acoustical differences could not explain the reported results. The A, F, and G elements were balanced so that they occurred equally often in each of the correct and violation sequences (Fig. 1B). Half of the violation and correct sequences were also balanced for the presence of the C and D elements, but it was difficult to achieve this balance in the other half of the sequences without introducing other potential confounds. Nonetheless, we found that acoustical differences cannot explain the results for the following reasons. First, macaque eye-tracking results by acoustical element (see Fig. 4; Results) showed that a comparable pattern of stronger responses to elements in violation versus correct sequences were made in response to all of the elements. Therefore, the macaques do not simply respond strongly to certain elements, but their responses vary based on the type of sequence in which they occurred (correct or violation). Second, an analysis of the average eye position in response to the C and D elements [ANOVA factors: element (C or D), condition (correct or violation) and monkey] showed the expected main effect of condition ($p = 0.001$) and monkey ($p = 0.008$), but no effect was observed for the element factor ($p = 0.13$) and no interactions were seen between the elements and condition or monkey (all p values >0.1). Therefore, the responses cannot be explained by a preference for any acoustical element but can be explained by the context in which the element occurs.

Participants: Rhesus macaques. Thirteen male Rhesus macaques (*Macaca mulatta*) participated in this experiment. The macaques were in two separate group-housed colonies. The animals were individually separated in these colonies for testing, wherever possible.

Participants: Common marmosets. Four common marmosets (*Callithrix jacchus*) participated in this experiment. The marmosets were in a single group-housed colony. The animals were individually separated in the colony for testing, wherever possible.

Habituation phase. During the habituation phase, the animals were presented with habituation sequences in a randomized order (Fig. 1B). The sequences were presented from a concealed audio speaker (rate of 9 sequences/min; intersequence interval = 4 s). Habituation occurred for 2 h on the afternoon before the experiment, when the animals were quiet and relaxed, but a few hours before the lights would be turned off for them to sleep. The following morning the animals were rehabituated to the sequences presented in a randomized order for 10 min, immediately before the start of the experiment.

Test phase. Video cameras were set up early in the morning to allow the animals to become habituated to their presence. During testing, a randomly selected test sequence of the eight (correct or violation) sequences (Fig. 1B) was individually presented (4 times each, for a total of 32 testing trials; at an average rate of 1/min; intersequence intervals ranged between 45 and 75 s). Each animal's orienting responses were video recorded for offline analysis (JVC and Sony digital video cameras; 720 × 576 resolution; 25 frames/s). To obtain sufficient results with the four marmosets that were available for testing, the animals were tested on four separate occasions, with at least 1 week separating the testing session. No differences were observed between monkeys or testing sessions, suggesting that the results could not be explained by any learning effect or individual differences (see Video-coding procedure, below).

Video-coding procedure. We refined the traditional video-coding procedure to minimize subjectivity in video-coding analysis. First, the audio track for each video was digitally scrambled so that it was not possible to identify the sequence condition. The videos from each animal were independently blind-coded by three raters (coauthors: A.E.M., H.S., and B.W.). Each rater coded orienting responses based on eye, head, and/or body movements in the direction of the concealed audio speaker that presented the stimulus sequences. The strength of the orienting responses were recorded on a five-point Likert scale, 1 = no orienting response; 2 = probably no response; 3 = ambiguous response; 4 = probable orienting response; 5 = definite orienting response. All analyses were based on trials on which a majority of raters (2 of 3) agreed that an

unambiguous response (strength ≥ 4) occurred. We analyzed the proportion of trials on which the animals unambiguously responded and the average duration of the orienting response in these trials (Fig. 2).

To understand the variability between the four marmosets and four testing sessions, these were included as factors within two repeated-measures ANOVAs (RM-ANOVAs; because of the limited degrees of freedom all factors could not be included in the same model). One RM-ANOVA modeled the between-subject "condition" (2 levels: correct vs violation sequences) factor and the within-subject factor of "session" (4 levels). The other analysis modeled the between-subject condition factor and the within-subject factor of "marmoset" (4 levels). These analyses all revealed a significant main effect of condition, but neither showed any effect of marmoset, session or interactions with condition (all p values >0.4). These additional results confirm that we observed stable performance between animals and across session, suggesting a homogeneous dataset was available for further analysis with no significant differences between animals or testing sessions.

Inter-rater reliability: macaques. Three raters coded all of the videos. Inter-rater reliability was calculated pairwise between the raters. In the macaque experiment, the raters, on average, had exact agreement on the strength of the response (on the five point scale) on 75.4% of the trials and were within one response point from each other on 85.1% of the trials. Also, Cohen's Kappa (Landis and Koch, 1977) revealed "substantial" agreement, $K = 0.67$. Only trials on which a majority of the raters agreed that an unambiguous response had occurred were included in further analyses. The macaques were rated as unambiguously responding to 14.7% of all recorded trials by a majority of raters resulting in 16 grammatical and 45 ungrammatical response trials (total of 61) used for analysis.

Inter-rater reliability: marmosets. Three raters coded all of the videos. Inter-rater reliability was calculated pairwise between the three raters. In the marmoset experiment, the raters had exact agreement on the strength of the response (on the five point scale) on 49.8% of the trials and were within one response point from each other on 80% of the trials. Cohen's Kappa (Landis and Koch, 1977) revealed "fair" to "moderate" agreement, $K = 0.39$. The marmosets were rated as unambiguously responding on 22.8% of all of the recorded trials by a majority of raters resulting in 60 grammatical and 57 ungrammatical response trials (total of 117) used for analysis. These numbers in comparison to those of macaques (above) indicate that the marmoset data were not statistically underpowered in relation to those that were available for analysis from the macaques. See Results for further details.

Eye-tracking experiment

Participants. Three adult male Rhesus macaques (*Macaca mulatta*) previously trained on a fixation task and acclimated to head immobilization.
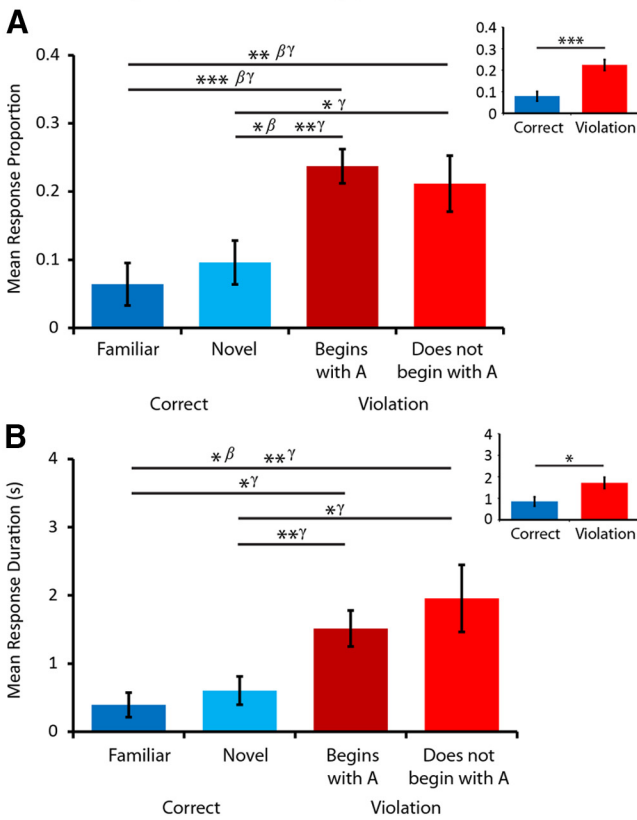
Stimuli. The stimulus sequences were identical to those used in the video-coding experiment (Fig. 1B).

Procedures. Animals were seated in a primate chair 60 cm in front of a computer monitor, displaying a fixation circle, and two audio speakers (Creative Inspire T10) horizontally positioned at ±30° visual angle (Fig. 3A). Following 25% of successful fixation trials, a stimulus sequence was presented from either the left or the right audio speaker, and eye tracking data were recorded (Fig. 3).

Habituation phase. During each habituation phase with each animal, one of seven sets of habituation sounds was randomly selected and played to the animal over both audio speakers for 30 min. Each of the habituation sets consisted of the nine habituation sequences presented in a randomized order (Fig. 1B); rate of presentation: 9 sequences/min; intersequence interval = 4 s.

Testing phase. Following the habituation phase was a testing run consisting of multiple trials. Each trial began when the animal engaged a red fixation spot in the center of the screen to center the eyes. If the animal continuously fixated for 2 s it was given a juice reward for fixating, and 25% of the successful fixation trials were followed by a testing trial in which a randomly selected testing sequence (of the 8 possible, see Fig. 1B) was randomly presented from either the right or the left audio speaker. The trials on which a testing sequence was presented were separated by on average four trials where no test sequence was presented and the animal only fixated. Eye-tracking data were
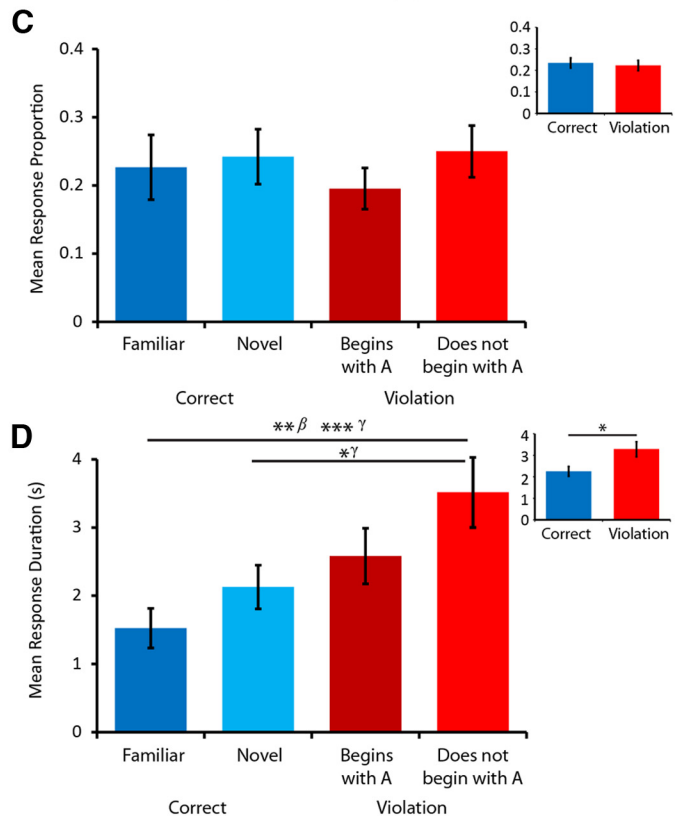
Figure 2. Video-coding experiment results in Rhesus macaques and common marmosets. (**A**, **C**) Mean proportion of trials (±SE across animals) on which the Rhesus macaques and common marmosets made unambiguous looking-responses as evaluated by a majority of 3 raters (see Materials and Methods). Subpanels indicate responses to correct and violation sequences, main panels display results to specific subsets of the correct or violation conditions. **B**, **D**, Mean response durations (±SE across animals) in macaques and marmosets in response to correct and violation sequences (subpanels) and to the four subcategories of stimuli. Both Bonferroni (β) and LSD (γ) *post hoc* tests are reported for all significant contrasts, *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

collected throughout the fixation and test sequence presentation periods (220 Hz infra-red eye tracker, Arrington Research). Experimental data were collected in 1–5 separate testing runs per day. Each testing run included at least eight trials (one presentation of each test sequence in a randomized order, see Fig. 1B). The animal was given a short break between each testing run, during which the animal listened to a new randomized set of habituation sequences for 5 min to rehabituate him to the AG structure. After this, another testing run began if the animal remained motivated to fixate to start each trial.

Eye-tracking experiment: data analysis. The three macaques participated in 25, 25, and 26 testing runs, respectively. Only the first eight trials of each testing run were used for further analysis since all of the animals completed these. The eye-tracking data for each trial contained both a 2 s baseline period during which the animal fixated on the central fixation spot and a subsequent period during which the test sequence was randomly presented from one of the two audio speakers. Significant-looking responses to the test sequences were defined individually for each animal as looks toward the presenting audio speaker (left or right) exceeding 3 SDs of the variability in the baseline eye fixation period. The analysis included the time from stimulus onset up to the point when the animal looked in the opposite direction for >200 ms. This identified when the animal seemed to lose interest in the test sequence and looked >3 SDs of baseline variability toward the opposite, silent audio speaker (Fig. 3B). The length of the response window for the three monkeys (M) was as follows: M1 = 2128 ms, M2 = 2984 ms, M3 = 4180 ms. The data were also analyzed using a fixed 3000 ms window and the pattern of results was comparable to those with the individually defined analysis windows. Within the response period, we analyzed durations of responses, defined as the proportion of time in the analysis window that the animal spent looking toward the presenting audio speaker, beyond 3SD of the baseline fixation period (Fig. 3B). We also analyzed the average eye deflections in the direction of the presenting speaker. For analysis of the average looking-response to individual elements, the window was the time during which the element was presented with an adjustment for how long, on average, it took the animal to breach the 3 SD criterion to look toward the presenting speaker at the start of the test sequence.

Eye-tracking experiment: analysis of specific violation sequences. To assess whether the macaques were sensitive to subtle, additional violations in later parts of the testing sequences, we analyzed whether the macaques were sensitive to differences between two sequences (see Fig. 5, i and ii), which begin identically but then differ in their number of violations later in the sequences. Mean difference plots between sequence i and ii were generated for each monkey (see Fig. 5), across the sequence repetitions (each sequence was repeated, respectively, 25, 25, and 26 times in macaques 1, 2, and 3). Then, 95% confidence intervals were generated using a bootstrapping procedure as follows. Within the early part of the sequence, during the presentation of the first two elements that are identical between the sequences, we created a data matrix of the eye-traces within this period (time) by the number of repeats of the two sequences. We then shuffled the sequence labels 1000 times to generate the null-hypothesis distribution of differences to determine the 5 and 95% confidence intervals (CI; see Fig. 5). Deviations of the difference in eye trace below the 5% CI reflect responses in favor of sequence i with the fewer violations; differences above the 95% CI would show a preference for sequence ii. Last, for any significant deflection below the 5% or above the 95% CI, we calculated the area (representing both the time and magnitude of the deviation across the CI) that breached this significance threshold.
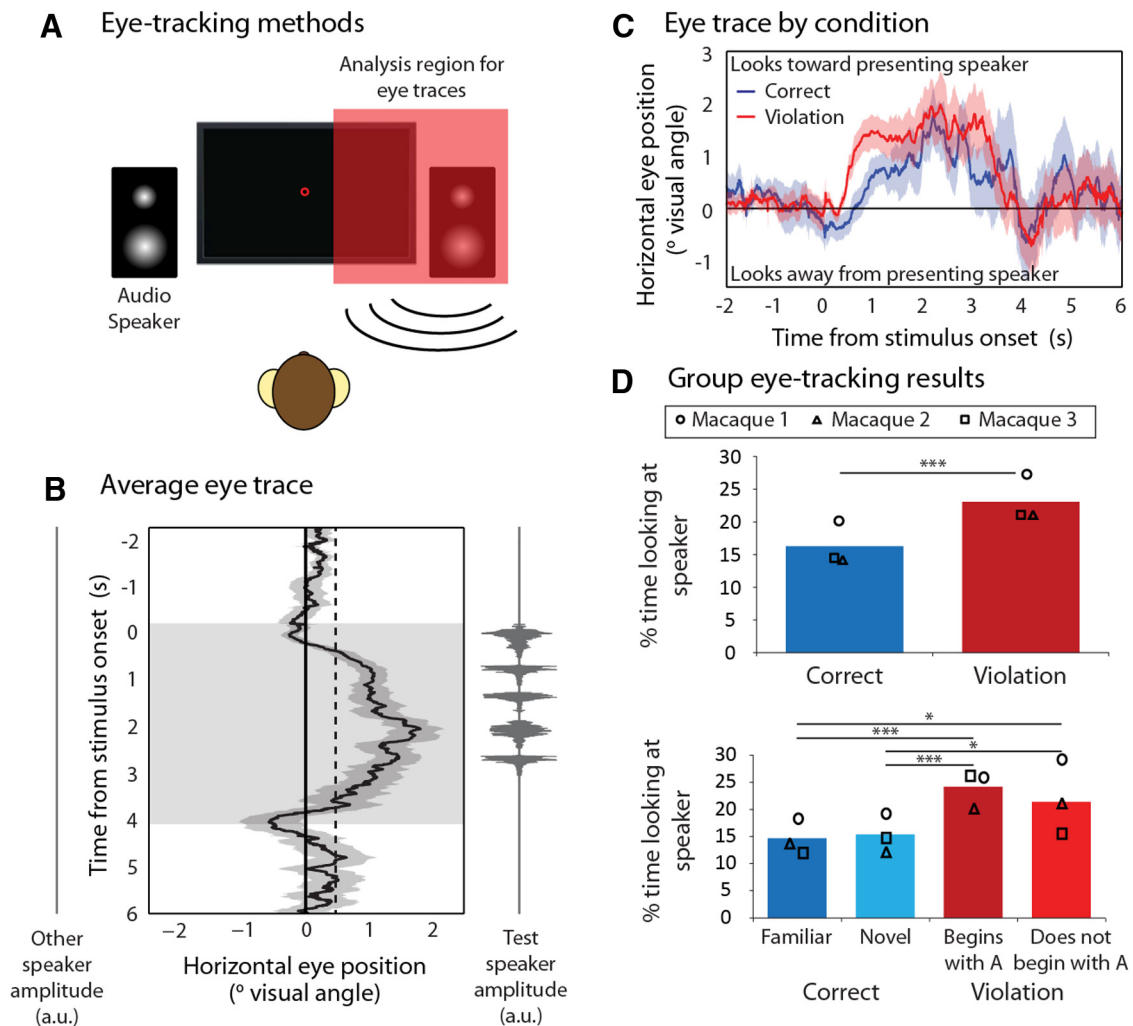
Figure 3. Eye-tracking measurement of preferential looking-responses to different testing sequences. ***A***, Schematic of macaque eye-tracking experiment. ***B***, Average eye trace from one monkey (±SE across trials). Positive values on the horizontal axis indicate eye movements toward the audio speaker (left or right) that presented a given test sequence. The dotted line denotes 3 SDs of the variance in eye position during fixation, which was used for analysis of significant looking-responses (shaded area is the individually defined response period; see Materials and Methods). ***C***, Mean eye traces (±SE) to the correct and violation sequences for the same monkey. ***D***, Group eye-tracking results including individual results by monkey: Top shows mean response duration (%) of looking-responses to the correct and violation conditions. Bottom shows results for the "familiar" and "novel" correct test sequences and violation sequences that (like the correct sequences) "begin with A" or those that "do not begin with A." *$p < 0.05$, ***$p < 0.001$. a.u., Arbitrary units.

Results

Video-coding experiments

After habituating the monkeys to exemplary sequences following the AG structure, we tested both macaques' and marmosets' orienting responses to well formed (correct) sequences that followed the AG structure, compared with sequences that violated the AG structure in certain ways. The animals' responses were videotaped. To minimize experimental bias in the analysis of videotaped responses, three raters blind coded all of the videos before analysis and only trials in which there was majority rater consensus were analyzed (see Materials and Methods).

The 13 macaques showed a significantly higher proportion of orienting responses to the violation sequences than the correct sequences (paired samples t test, $t_{(12)} = 7.898$, $p < 0.001$; Fig. 2A, subpanel). We analyzed the two different types of correct and violation sequences to clarify whether the observed effect depends on simpler strategies such as familiarity and/or the animals only noticing violations in the first sequence element (i.e., sequences that, unlike the correct sequences, "do not begin with A," Fig. 1B). We used an RM-ANOVA with four levels of the se-

quence condition factor: "familiar," "novel," "begin with A," and "do not begin with A" (Fig. 2A). Within the main effect for sequence condition ($F_{(3,36)} = 9.146$, $p < 0.001$; Fig. 2A), Bonferroni comparisons showed differences between several key contrasts, including between "novel" correct sequences and violation sequences that "begin with A," i.e., those that cannot be identified by either familiarity or an unexpected initial element (Bonferroni corrected, $p = 0.03$; Fig. 2A). No differences were observed between "familiar" and "novel" correct sequences ($p = 1.0$) or violation sequences that "begin with A" or "do not begin with A" ($p = 1.0$). A similar pattern of effects was observed when analyzing the duration of responses (correct and violation sequences: $t_{(12)} = 2.330$, $p = 0.038$; RM-ANOVA "condition" factor with 4 levels: $F_{(3,36)} = 5.276$, $p = 0.004$; Fig. 2B). These results together suggest that not only do the macaques respond to violations of the AG, but also that their responses cannot be attributed only to superficial differences between the sequences, such as novelty or monitoring only the initial parts of the sequences.

Four marmosets were available for study, thus, to obtain sufficient data for analysis they were each tested four times. Each

testing run was separated by at least 1 week and followed an identical procedure to the macaque experiment, including a habituation and testing phase. First, to investigate whether there were differences between monkeys or testing sessions, these were entered as factors into RM-ANOVA models (see Materials and Methods, Video coding experiment data analysis). There were no effects of monkey or session in any of the analyses and these factors did not interact with the experimental effects (all p values > 0.4), suggesting that the testing sessions were independent and homogenous (i.e., there were no strong across-session learning effects). The marmosets did not discriminate between the different conditions based on the frequency of looking responses ($p > 0.5$, Fig. 2C) but did based on the duration of their looking responses, which were significantly longer for the violation than the correct sequences (correct vs violation sequences: $t_{(12)} = 2.142$, $p = 0.043$; RM-ANOVA "condition" factor with 4 levels: $F_{(1,12)} = 5.895$, $p = 0.032$; Fig. 2D). However, unlike the macaque results, Bonferroni comparisons only showed an effect between the "familiar" sequences and those that "do not begin with A" ($p = 0.003$). Even with a less conservative LSD correction for multiple comparisons the only additional effect observed was between "novel" and "do not begin with A" ($p = 0.035$, Fig. 2D), therefore the marmosets' responses appear to be based primarily on familiarity or noticing violations at the beginning of the sequences. Interestingly, overall, the marmosets responded more strongly than macaques (compare the four subpanels in Fig. 2A–D), both in terms of the proportion of responded trials (main effect of species, $F_{(1,23)} = 6.611$, $p = 0.017$; main effect of condition, $F_{(1,23)} = 22.963$, $p < 0.001$; significant interaction, $F_{(1,23)} = 12.869$, $p = 0.002$) and the duration of responses (main effect of species, $F_{(1,23)} = 22.162$, $p < 0.001$; main effect of condition, $F_{1,23} = 9.449$, $p = 0.005$; no interaction, $p = 0.656$). These observations and the number of trials available for analysis (see Materials and Methods) suggest that the marmoset results do not appear to have been statistically underpowered relative to the macaque results.

In summary, the video-coding results show that the marmosets respond for longer durations to sequences that violate the AG structure. However, their results appear to stem primarily from sensitivity to violation sequences that "do not begin with A" (creating a simple violation in the first position of the sequence). This is shown in the duration of marmoset orienting responses by a significant difference between "familiar" or "novel" sequences and those violation sequences that "do no begin with A." Rhesus macaques, like the marmosets, responded for longer durations to the violation compared with correct sequences. However, the macaques showed stronger responses to violation sequences that "begin with A" compared with "novel" correct sequences, whereas this effect was not observed in the marmosets. These observations reveal that the macaques' results cannot be based on the familiarity of the sequence or violations in the initial element position. Eye-tracking experiments were conducted with three of these macaques to further probe their sensitivity to the AG structure and to investigate responses at later positions in the sequence. Unfortunately, eye tracking with the smaller marmosets was not technically possible for this study.

Eye-tracking experiment
In three macaques, we asked whether infrared eye-tracking measurements would reveal differential looking-responses between the correct and violation sequences. The approach is shown in Figure 3, *A* and *B*, and Materials and Methods. The results from the three Rhesus macaques were analyzed using an RM-ANOVA
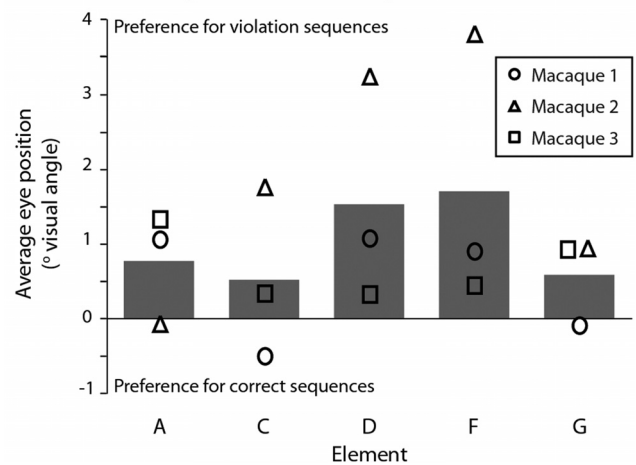

Figure 4. Eye-tracking of looking-responses to individual elements. Group (and individual) mean difference plot of responses to "violation" minus "correct" sequences in response to each of the five stimulus elements (A, C, D, F, and G). Positive numbers reflect stronger looks to violation sequences than to correct sequences.

with two factors: "monkey" and "sequence condition" (levels: "correct" and "violation"). The results confirmed those seen in the video coding experiment: the animals made significantly longer looking-responses to the violation sequences (significant main effect of sequence condition: $F_{(2,73)} = 20.297$, $p < 0.001$; Fig. 3D). Although individual animals differed in their looking times toward the presenting audio speaker (significant main effect of monkey, $F_{(2,73)} = 4.055$, $p = 0.021$), there was no interaction between sequence condition and monkey factors ($p = 1.0$). Moreover, the eye-tracking approach revealed longer look durations to violation than correct sequences in the individual macaques ($t_{(24)} = 3.137$, $p = 0.004$; $t_{(24)} = 3.129$, $p = 0.005$; and $t_{(25)} = 2.023$, $p = 0.05$, respectively).

An RM-ANOVA with four levels of the sequence condition factor: "familiar," "novel," "begin with A," and "do not begin with A" (Fig. 3D) showed a significant main effect for sequence condition ($F_{(3,219)} = 10.057$, $p < 0.001$). Bonferroni comparisons revealed significant differences between: (1) "novel" and "begin with A" ($p = 0.001$); (2) "novel" and "do not begin with A" ($p = 0.014$); (3) "familiar" and "begin with A" ($p < 0.001$); and, (4) "familiar" and "do not begin with A" ($p = 0.012$); see Figure 3D. There was no significant difference between responses to familiar versus novel correct sequences ($p = 1.0$), nor any effect between responses to the violation sequences that "begin with A" versus "do not begin with A" ($p = 1.0$). Furthermore, there was no interaction between the condition and monkey factors ($p = 1.0$). These effects were confirmed at the individual level (main effect of condition $F_{(3,72)} = 3.715$, $p = 0.015$; $F_{(3,72)} = 4.745$, $p = 0.004$; $F_{(3,75)} = 5.08$, $p = 0.003$, respectively). These results recapitulate the video-coding results and suggest that the macaques' abilities to discriminate correct from violation responses do not depend on sequence familiarity or on rote memorization during the habituation phase. Given that the monkeys seem to monitor the sequences for violations after the first element, we asked whether they could also monitor the rest of the sequence for possible violations. In particular, we assessed if they responded to violations beyond the second position in the violation sequences (at which point the branching structure of the AG becomes more evident, Fig. 1B).

To better determine the extent of macaque AG learning abilities, we first compared eye movements in response to identical

acoustical elements, within the context of either correct or violation sequences (Fig. 4). This was done by performing an ANOVA with the factors of "sequence condition" ("correct" or "violation"), "element" (A, C, D, F, or G), and "monkey" (3 levels). Critically, the main effect of sequence condition ($F_{(1,73)} = 11.978$, $p = 0.001$; Fig. 4) did not interact with element ($p = 0.1$), nor was there a main effect of element itself ($p = 0.6$). Thus, the stronger looks to violation sequences cannot be explained by the animals' responses to any individual element. Furthermore, there was no correlation between the magnitude of looks to different elements and the position in the correct or violation sequence where the element occurred (correct: $r = -0.01$, $p = 0.8$; violation: $r = -0.05$, $p = 0.3$), suggesting that the animals' responses to the violations were not based on the preferences for any specific elements, or due to increased responsiveness at a specific time throughout the sequences.

We also directly investigated eye movements at particular positions in the sequences, focusing the analysis on two specific violation sequences (Fig. 5). These sequences begin identically, have their initial violation in the second position, and contain the same elements in positions 3–5. However, the elements in positions 3–5 have a different order, which generates an additional violation in one of the sequences at the third element (see sequence *ii*, Fig. 5A). We asked whether the animals were sensitive to this later difference between the two violation sequences. A bootstrapped statistical analysis of the animals' eye-traces demonstrated that two of the macaques showed strong significant responses in favor of sequence *ii*, containing the additional violation, which resulted in an area above the significance threshold at least a factor of 3 greater than any such preference seen either for sequence *i* or during sequence positions 1–2 where the two sequences are identical. No difference could be observed between the sequences in macaque 3 (Fig. 5). Moreover, the variability in looks in macaques 1 and 2 to these violation sequences (both of which start with "AF") do not support the notion that a special interest in the second element "F" alone captures the animals' attention and results in persistent looking responses to all subsequent elements in the violation sequences (Fig. 6). These results suggest that a significant sensitivity to a subtle violation later in the sequences can be measured in a majority of the three animals studied.

Discussion

New World monkeys (marmosets) were able to notice simple violations of the AG, such as those in the first sequence element.
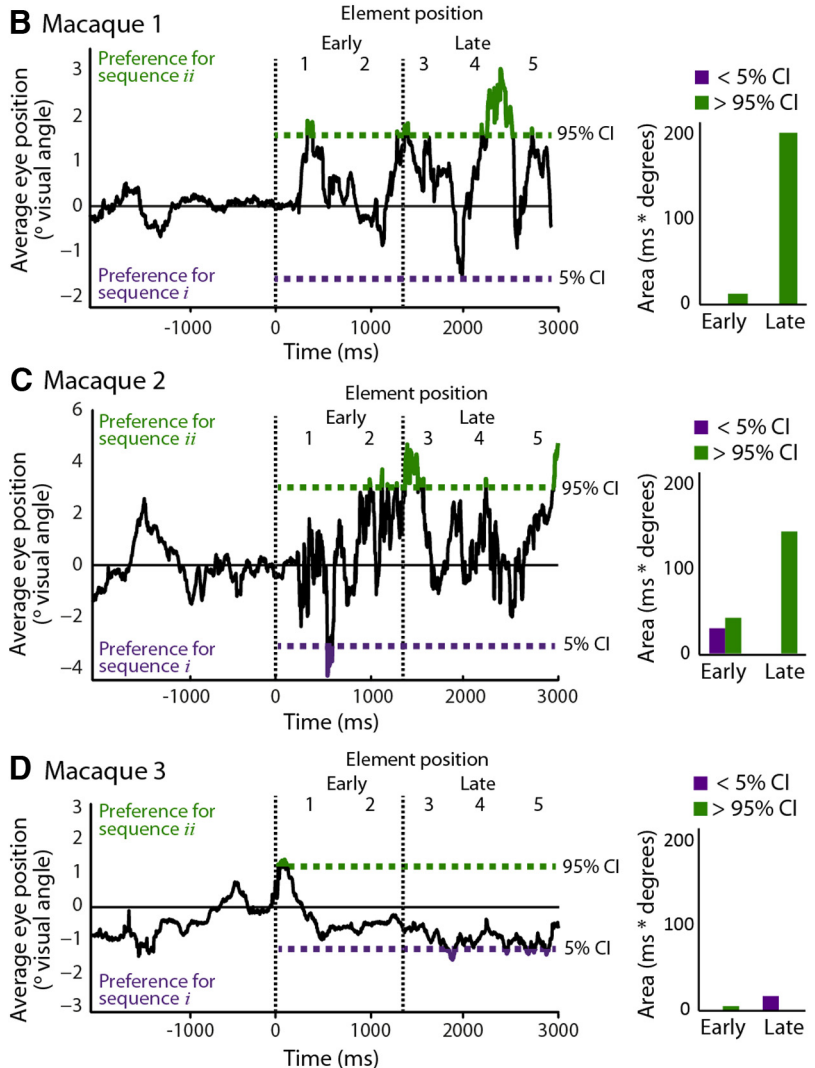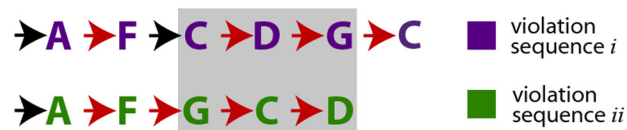

Figure 5. Individual macaque eye-tracking sensitivity to violations at specific sequence positions (difference plots). **A**, Schematic plot of two of the violation sequences, identifying legal transitions (black arrows) and violations (red arrows). Violation sequence *ii* (green) contains one more violation than sequence *i* (purple) in the transition between the second and third elements in the sequence. **B–D**, Average difference in looking preferences toward sequence *ii* (positive numbers) or sequence *i* (negative numbers), across the repetitions of each sequence; respectively: 25, 25, and 26 for each animal. Vertical black lines denote stimulus onset (at 0 ms) and the onset of element 3, where the sequences diverge. Dashed lines indicate CI (based on bootstrapped differences, 1000 permutations, see Materials and Methods). Also shown are the areas >95% or <5% CI (bar plots, right) where each animal made statistically significant looks in favor of either sequence.

In contrast, Old World monkeys showed evidence for the capacity to learn more of the nondeterministic components of the AG throughout the sequences. We consider the results relative to other studies and the relationships between the studied AG and language and song structure.

AGL results in relation to previous studies

Many animals are able to recognize a single element from a set, e.g., recognizing a vocalization from a set of vocalizations
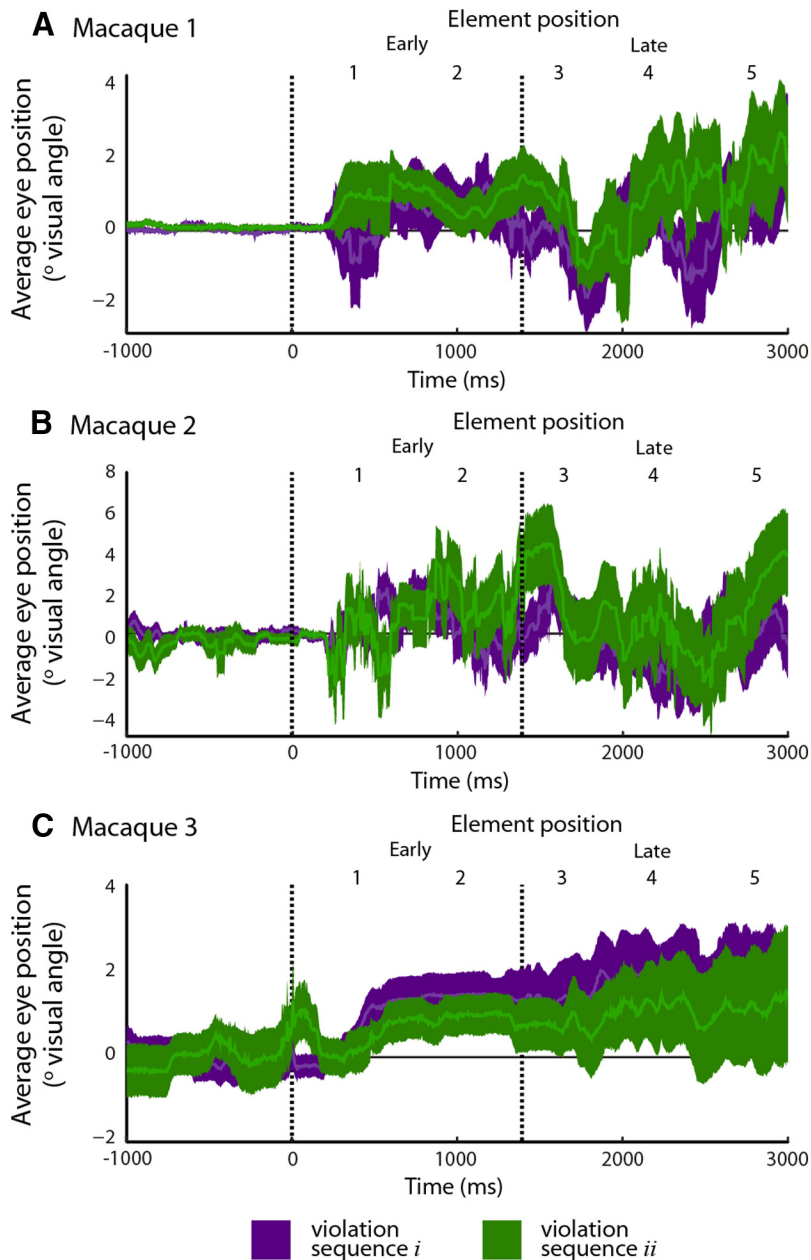
Figure 6. Individual macaque eye traces in responses to specific violation sequences. Average eye position (degrees visual angle ± SE) in response to violation sequence (*i*) and (*ii*) for the three macaques. ***A–C***, These sequences are identical for the first two elements but sequence *ii* then contains an additional violation before the start of element 3, relative to sequence *i* (Fig. 5*A*). Vertical black lines denote stimulus onset (at 0 ms) and the onset of element 3. Stronger responses to violation sequence (*ii*) can be seen in macaques 1 and 2 (***A, B***) after the onset of the third element (note the areas of separation between the colored areas of the SEs for the two sequences) but not for macaque 3 (***C***) who may only show a slight preference for sequence *i* during the period when both sequences are identical (element positions 1–2).

(Hauser et al., 2001). Relatedly, Newport et al. (2004) suggest that tamarins can learn the relationship between the first and last syllable in a triplet sequence with more variable intervening elements.

Other studies that have tested nonhuman animals with more complex AG structures have produced more variable results. For instance, evidence for $A^{n}B^{n}$ structure learning has only been obtained in birds (Fitch and Hauser, 2004; Gentner et al., 2006; van Heijningen et al., 2009; Stobbe et al., 2012). The $A^{n}B^{n}$ structure is of interest because it was designed to contain nonadjacent associations between different pairs of stimuli (Hauser et al., 2002; Bahlmann et al., 2008). However, it remains unclear whether any nonhuman animal can naturally produce or learn nonadjacent hierarchical relationships such as those found in human language (Berwick et al., 2011; Hurford, 2012; Jäger and Rogers, 2012; Petersson and Hagoort, 2012). The quantitative parameter space in this study, within which AG structures and animal AGL studies were evaluated (Fig. 1*A*), sidesteps this controversy and highlights other interesting aspects of the complexity of animal AGL that remain under-studied.

The results reported here show that both common marmosets and Rhesus macaques are sensitive to violations of a forward branching, nondeterministic AG. In macaques, the results rule out trivial explanations based on familiarity, acoustic differences between well formed (correct) and ill-formed (violation) sequences, and (at least in 2 of 3 of the macaques tested) monitoring only the initial part of the sequences (Figs. 3*C,D*, 5, 6). However, the video-coding results suggest that marmosets' responses could be interpreted on the basis of familiarity or on noticing only violations in the first position in a sequence. It seems unlikely that the marmoset results would have differed even if more marmosets had been available for testing or if the eye-tracking techniques had been feasible with them, for the following reasons. The four marmosets were tested over four testing sessions (each separated by at least 1 week), yet even with the additional testing sessions they showed no evidence of adopting a more complex AGL strategy. Furthermore, the marmosets were more responsive than the macaques, so their results cannot be attributed to reduced statistical power. Also, the results of both the macaque video coding and eye tracking results were complementary; therefore, it is unlikely that an eye-tracking experiment would have yielded very different results to the video coding experiment in marmosets. For instance, even macaque 3 who, unlike macaques 1 and 2, showed no significant sensitivity in the eye tracking experiment to the subtle violation in a later part of one of the violation se-

(Moore, 2004). An increase in sequence learning complexity, of which all animals tested appear capable, involves learning the relationship between two different elements within a sequence, e.g., recognizing adjacent relationships between A and B stimulus classes in $(AB)^{n}$ structures (Fig 1*A*). Another facet of sequencing relationships is sensitivity to the transitional probabilities between pairs of elements in a sequence (Saffran et al., 1996; Saffran et al., 1999). Tamarin monkeys were reported to have shown stronger dishabituation responses to sequences of three syllable triplets that contained low probability transitions between the syllables, transitions which the monkeys had rarely encountered

quences (Fig. 5), nonetheless noticed violations in sequences that begin with A, which start identically to the well formed (correct) sequences (Fig. 3D). This important observation is seen in the macaque, but not in the marmoset, video-coding results (Fig. 2). Interestingly, our observations of the marmoset behavior correspond to those from another AGL study in New World monkeys (cotton-top tamarins), whereby the tamarins only showed significant familiarization-based learning effects when the same AG sequences that were used for habituation were also used for testing (Saffran et al., 2008). Notably, in that study, human infants readily learned the AGL structure (including under less predictable conditions) and their responses generalized to novel correct sequences, the latter of which we only see in the macaque results.

It is of course possible that under different experimental conditions—such as with operant conditioning, by using different stimuli as elements in the AG [such as tones (Saffran et al., 1999)], or with more exposure to the sequences (Miles and Meyer, 1956)—marmoset and tamarin monkeys might be shown to be capable of more comprehensive learning of nondeterministic AGs or even more complex relationships in AG structures. However, the current results motivate the hypothesis that, under comparable experimental conditions, species more closely evolutionarily related to humans or to vocal learners such as songbirds, might have a relative advantage in the complexity of the AG structures that they are able to learn over more distantly related species. Such a hypothesis requires further testing with many more species, which might refine, refute, or support it.

Distinction between vocal production and auditory learning capacities

It might seem surprising that macaques show evidence for deeper AGL than marmosets, given that marmosets are more vocal, and in our study responded more frequently than the macaques. However, it is important to distinguish between vocal production and auditory learning since these capacities seem to be subserved by different neurobiological pathways and mechanisms (Jarvis, 2004; Petkov and Jarvis, 2012). Regarding vocal production, evidence of combinatorial calling has been reported in some Old World monkeys [e.g., putty-nosed monkeys, *Cercopithecus nictitans* (Arnold and Zuberbühler, 2006)], but whether macaques or marmosets are also capable of this is currently unknown. Rather than vocal communication, the sequence-structure learning abilities that we tapped into with these implicit AGL tasks may relate to learning processes. For example, earlier studies have suggested that macaques were able to learn a discrimination-learning task, including with a delay, more quickly than marmosets (Miles and Meyer, 1956; Miles, 1957). Furthermore, it is possible that many nonhuman primates can learn aspects of AG structures because they can evaluate patterns in sensory input [or the structure of social interactions (Bergman et al., 2003); like the movement patterns of others (Schmitt, 2010)]. However, our understanding of these abilities would benefit from more formal structural analysis and direct cross-species comparisons, such as those presented here.

Relationship of the current AG structure to natural song and language structures

Relative to many commonly used AGs, this paradigm departs from the requirement that stimuli are categorized into only two stimulus classes. Rather, several elements, both obligatory and optional, contribute to the structure, as exemplified by the original AGL study in humans by Reber (1967). A number of AGs used to test humans have a forward-branching structure similar to that used here (Reber, 1967; Friederici et al., 2002; Uddén et al., 2008). Branching structures with varying levels of predictability or linearity can also be observed in the natural song production of several species. For instance, zebra finches produce a relatively linear song (Honda and Okanoya, 1999), while the songs of other birds (Okanoya, 2004) and even some whales (Hurford, 2012) show more branching transitions, which form phonological "syntax-like" structures of interest to linguists and other scientists (Bolhuis et al., 2010; Berwick et al., 2011).

Word transitions in sentences of natural languages are characterized by nondeterminism: sentences are not fixed, predetermined sequences, but vary considerably in composition, word transitions, and length. Well-formed sentences contain obligatory components (e.g., a subject and a finite verb in English declaratives), as well as varying numbers of optional categories (adjectives, adverbs, etc.), the positions of which depend on the other words in the sentence. Language learners must deal with unpredictable variation (Kam and Newport, 2009) and appear to have a general bias to reduce such variation during learning (Smith and Wonnacott, 2010). Thus, the capacity to evaluate hierarchical relationships of the types present in human language may need to be accompanied by processes that allow us to cope with sequence variability, which are capacities that appear to have clearer evolutionary origins.

Conclusions

We report evidence of a novel level of AGL complexity in Old World monkeys (Rhesus macaques). The macaque results cannot easily be attributed to simple strategies such as responding to acoustic differences, the novelty of sequences, or only recognizing violations early in the sequences. While the common marmosets (New World monkeys) also showed dishabituation responses to violations of the AG structure, the results failed to rule out a reliance on simple strategies. Such behavioral results provide part of the initial foundation required for neuronal level investigations of different aspects of syntactic precursors in primate laboratory model systems such as marmoset and macaque monkeys.

References

Abe K, Watanabe D (2011) Songbirds possess the spontaneous ability to discriminate syntactic rules. Nat Neurosci 14:1067–1074. CrossRef Medline

Arnold K, Zuberbühler K (2006) Language evolution: semantic combinations in primate calls. Nature 441:303. CrossRef Medline

Bahlmann J, Schubotz RI, Friederici AD (2008) Hierarchical artificial grammar processing engages Broca's area. Neuroimage 42:525–534. CrossRef Medline

Beckers GJ, Bolhuis JJ, Okanoya K, Berwick RC (2012) Birdsong neurolinguistics: songbird context-free grammar claim is premature. Neuroreport 23:139–145. CrossRef Medline

Bennett CL, Davis RT, Miller JM (1983) Demonstration of presbycusis across repeated measures in a nonhuman primate species. Behav Neurosci 97:602–607. CrossRef Medline

Bergman TJ, Beehner JC, Cheney DL, Seyfarth RM (2003) Hierarchical classification by rank and kinship in baboons. Science 302:1234–1236. CrossRef Medline

Berwick RC, Okanoya K, Beckers GJ, Bolhuis JJ (2011) Songs to syntax: the linguistics of birdsong. Trends Cogn Sci 15:113–121. CrossRef Medline

Bickerton D, Szathmary E (2009) Biological foundations and origin of syntax. Cambridge, MA: MIT.

Bolhuis JJ, Okanoya K, Scharff C (2010) Twitter evolution: converging mechanisms in birdsong and human speech. Nat Rev Neurosci 11:747–759. CrossRef Medline

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

de Vries M, Christiansen MH, Petersson KM (2011) Learning recursion: multiple nested and crossed dependencies. Biolinguistics 5:010–035.

Fitch WT, Hauser MD (2004) Computational constraints on syntactic processing in a nonhuman primate. Science 303:377–380. CrossRef Medline

Friederici AD, Steinhauer K, Pfeifer E (2002) Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. Proc Natl Acad Sci U S A 99:529–534. CrossRef Medline

Friederici AD, Bahlmann J, Heim S, Schubotz RI, Anwander A (2006) The brain differentiates human and nonhuman grammars: functional localization and structural connectivity. Proc Natl Acad Sci U S A 103:2458–2463. CrossRef Medline

Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC (2006) Recursive syntactic pattern learning by songbirds. Nature 440:1204–1207. CrossRef Medline

Hauser MD, Glynn D (2009) Can free-ranging rhesus monkeys (*Macaca mulatta*) extract artificially created rules comprised of natural vocalizations? J Comp Psychol 123:161–167. CrossRef Medline

Hauser MD, Newport EL, Aslin RN (2001) Segmentation of the speech stream in a nonhuman primate: statistical learning in cotton-top tamarins. Cognition 78:B53–64. CrossRef Medline

Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: what is it, who has it, and how did it evolve? Science 298:1569–1579. CrossRef Medline

Honda E, Okanoya K (1999) Acoustical and syntactical comparisons between songs of the white-backed Munia (*Lonchura striata*) and its domesticated strain, the Bengalese finch (*Lonchura striata var. domestica*). Zool Science 16:319–326. CrossRef

Kam CL, Newport EL (2009) Getting it right by getting it wrong: when learners change languages. Cogn Psychol 59:30–66. CrossRef Medline

Hurford JR (2012) The origins of grammar: language in the light of evolution. Oxford: Oxford UP.

Jäger G, Rogers J (2012) Formal language theory: Refining the Chomsky hierarchy. Philos Trans R Soc Lond B Biol Sci 367:1956–1970. CrossRef Medline

Jarvis ED (2004) Learned birdsong and the neurobiology of human language. Ann N Y Acad Sci 1016:749–777. CrossRef Medline

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174. CrossRef Medline

Lonsbury-Martin BL, Martin GK (1981) Effects of moderately intense sound on auditory sensitivity in rhesus monkeys: behavioral and neural observations. J Neurophysiol 46:563–586. Medline

Miles RC (1957) Delayed-response learning in the marmoet and the macaque. J Comp Physiol Psychol 50:352–355. CrossRef Medline

Miles RC, Meyer DC (1956) Learning sets in marmosets. J Comp Physiol Psychol 49:212–222. CrossRef

Moore BR (2004) The evolution of learning. Biol Rev Camb Philos Soc 79:301–335. CrossRef Medline

Murphy RA, Mondragón E, Murphy VA (2008) Rule learning by rats. Science 319:1849–1851. CrossRef Medline

Newport EL, Hauser MD, Spaepen G, Aslin RN (2004) Learning at a distance II. Statistical learning of nonadjacent dependencies in a nonhuman primate. Cogn Psychol 49:85–117. CrossRef Medline

Okanoya K (2004) The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. Ann N Y Acad Sci 1016:724–735. CrossRef Medline

Petersson KM, Hagoort P (2012) The neurobiology of syntax: beyond string sets. Philos Trans R Soc Lond B Biol Sci 367:1971–1983. CrossRef Medline

Petersson KM, Folia V, Hagoort P (2012) What artificial grammar learning reveals about the neurobiology of syntax. Brain Lang 120:83–95. CrossRef Medline

Petkov CI, Jarvis ED (2012) Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. Front Evol Neurosci 4:12. Medline

Petkov CI, Wilson B (2012) On the pursuit of the brain network for protosyntactic learning in nonhuman primates: conceptual issues and neurobiological hypotheses. Philos Trans R Soc Lond B Biol Sci 367:2077–2088. CrossRef Medline

Pfingst BE, Hienz R, Miller J (1975) Reaction-time procedure for measurement of hearing. II. Threshold functions. J Acoust Soc Am 57:431–436. CrossRef Medline

Pfingst BE, Laycock J, Flammino F, Lonsbury-Martin B, Martin G (1978) Pure-tone thresholds for rhesus-monkey. Hear Res 1:43–47. CrossRef Medline

Reber AS (1967) Implicit learning of artificial grammars. J Verb Learn Verb Behav 6:855–863. CrossRef

Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. Science 274:1926–1928. CrossRef Medline

Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Cognition 70:27–52. CrossRef Medline

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. CrossRef Medline

Schmitt D (2010) Primate locomotor evolution: Biomechanical studies of primate locomotion and their implications for understanding primate neuroethology. In: Primate neuroethology (Ghazanfar AA, Platt ML, eds), pp 10–30. Oxford: Oxford UP.

Seiden HR (1958) Auditory acuity of the marmoset monkey (*Hapale jacchus*). Doctoral dissertation, Princeton University.

Seyfarth RM, Cheney DL, Marler P (1980) Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. Science 210:801–803. CrossRef Medline

Smith K, Wonnacott E (2010) Eliminating unpredictable variation through iterated learning. Cognition 116:444–449. CrossRef Medline

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 Lond B Biol Sci 367:1995–2006. CrossRef Medline

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. Science 292:290–293. CrossRef Medline

Uddén J, Folia V, Forkstam C, Ingvar M, Fernandez G, Overeem S, van Elswijk G, Hagoort P, Petersson KM (2008) The inferior frontal cortex in artificial syntax processing: an rTMS study. Brain Res 1224:69–78. CrossRef Medline

Uddén J, Ingvar M, Hagoort P, Petersson KM (2012) Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: investigating the push-down stack model. Cogn Sci 36:1078–1101. CrossRef Medline

van Heijningen CA, de Visser J, Zuidema W, ten Cate C (2009) Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. Proc Natl Acad Sci U S A 106:20538–20543. CrossRef Medline