

Nonadjacent dependency processing in monkeys, apes, and humans

Stuart K. Watson^{1,2*}, Judith M. Burkart³, Steven J. Schapiro^{4,5}, Susan P. Lambeth⁴,
Jutta L. Mueller^{6,7†}, Simon W. Townsend^{1,2,8†}

The ability to track syntactic relationships between words, particularly over distances (“nonadjacent dependencies”), is a critical faculty underpinning human language, although its evolutionary origins remain poorly understood. While some monkey species are reported to process auditory nonadjacent dependencies, comparative data from apes are missing, complicating inferences regarding shared ancestry. Here, we examined nonadjacent dependency processing in common marmosets, chimpanzees, and humans using “artificial grammars”: strings of arbitrary acoustic stimuli composed of adjacent (nonhumans) or nonadjacent (all species) dependencies. Individuals from each species (i) generalized the grammars to novel stimuli and (ii) detected grammatical violations, indicating that they processed the dependencies between constituent elements. Furthermore, there was no difference between marmosets and chimpanzees in their sensitivity to nonadjacent dependencies. These notable similarities between monkeys, apes, and humans indicate that nonadjacent dependency processing, a crucial cognitive facilitator of language, is an ancestral trait that evolved at least ~40 million years before language itself.

Copyright © 2020
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

INTRODUCTION

A standout, universal feature of human language is our ability to monitor syntactic relationships between words and phrases over distances, otherwise known as nonadjacent dependencies (“Non-ADs”; as opposed to “adjacent dependencies”—“ADs”) (1, 2). For example, in the English phrase /the dog that bit the cat ran away/, “ran away” is parsed as dependent on “the dog” and not “the cat.” Moreover, multiple Non-ADs can exist in a single sentence, such as when recursively embedding further words or tracking multiple agreement relations, e.g., for gender, tense, and number. Despite the critical importance of Non-AD processing for the comprehension and learning of language (2), its evolutionary origins remain ambiguous.

An emerging body of data from “artificial grammar” experiments, where subjects must compute predictive relationships between elements in strings of stimuli organized according to grammatical rules (3, 4), suggests that some species of monkey have the capacity to process Non-ADs in both auditory (5–8) and visual modalities (9–12). This has been argued to indicate an evolutionary continuity of the human capacity for Non-AD processing, traceable perhaps as far back as the last common ancestor of anthropoid primates (6). However, given the phylogenetic distance between New World monkeys and humans (approximately 40 million years ago), it is currently unclear whether this capacity is really ancestral or a product of convergent evolution. Compelling evidence for Non-AD processing in birds (13, 14) indicates that a convergent evolutionary scenario is plausible. Data from great apes are therefore needed to disentangle these competing hypotheses by shedding light on the capacity for Non-AD

processing in our more recent evolutionary past. While previous experimental data indicate that chimpanzees are able to process Non-ADs in visual patterns (15, 16), evidence of auditory Non-AD processing is notably absent. This is problematic, because it is unclear to what extent processing spatial patterns in a static visual image is related to the sequential processing of acoustic stimuli. Acoustic paradigms are arguably more pertinent to language processing (17), which primarily occurs in the auditory modality. Furthermore, there is evidence that suggests that these abilities do not directly map onto one another. For example, data from human statistical learning experiments (of which artificial grammar paradigms are a subset) show no correlation in learning performance across modalities nor are rules easily generalized between them, indicating some degree of domain specificity [reviewed by Frost *et al.* (18)]. The lack of evidence for auditory Non-AD processing in great apes thus severely complicates reconstructing the evolutionary history of this trait, wherein it is essential to examine “like-for-like” abilities across species. A gold standard for the comparative approach hence involves collecting comparable datasets through the cross-species application of matching experimental paradigms. This strategy has been extremely productive when examining other traits thought to be key to human cognition such as cumulative culture (19), hyper-cooperation (20), working memory (21), and theory of mind (22, 23). Here, we undertook such a directly comparative investigation of the evolutionary roots of Non-AD processing by implementing a standardized, auditory artificial grammar learning paradigm in common marmosets (*Callithrix jacchus*), chimpanzees (*Pan troglodytes*), and humans.

We familiarized individuals from each species to sets of artificial grammars: strings of arbitrary, frequency-modulated sine tones composed of either ADs (nonhuman primates) or Non-ADs (all species). Our grammars were combinations of two (AD condition) or three (Non-AD condition) elements from six categories of sound type (A, B, C, D, X1, and X2; Fig. 1), each of which consisted of 16 pitch-shifted variants. To determine whether nonhuman animals could process the grammars, we measured the time spent looking toward the speaker (24) after three types of probe sequence. These were (i) familiar sequences (“FS”), (ii) category combinations consistent with the

¹Department of Comparative Language Science, University of Zurich, Zurich, Switzerland. ²Center for the Interdisciplinary Study of Language Evolution, Zurich, Switzerland. ³Anthropological Institute and Museum, University of Zurich, Zurich, Switzerland. ⁴UT MD Anderson Cancer Research Center, Bastrop, TX, USA.

⁵Department of Experimental Medicine, University of Copenhagen, Copenhagen, Denmark. ⁶Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany.

⁷Department of Linguistics, University of Vienna, Vienna, Austria. ⁸Department of Psychology, University of Warwick, Coventry, UK.

*Corresponding author. Email: swatso88@gmail.com

†These authors contributed equally as co-last authors.

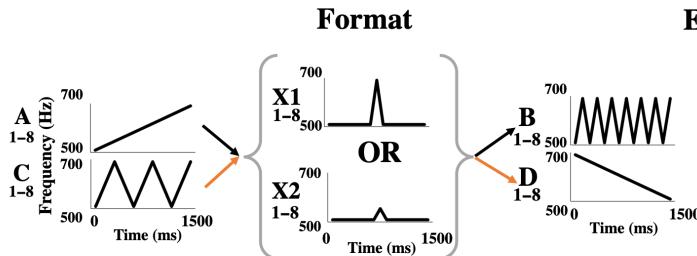
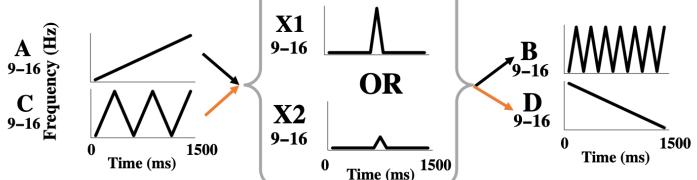
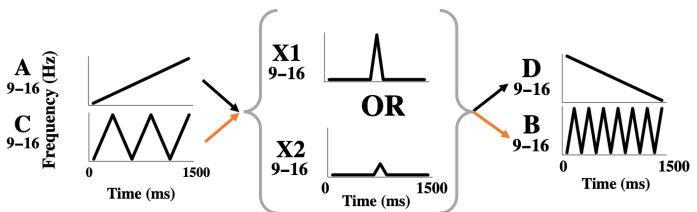
Sequence type**Familiarization****Generalization****Violation**

Fig. 1. Visual representation of each element and possible transitions between them for each sequence type (bracketed elements apply only to Non-AD condition), with examples. Numbers below category label indicate possible pitch variants. Y axis values in Familiarization row refer to pitch variant 1 of each element. Arrow color represents sequence transitions corresponding to a grammar (AXB and CXD) (see table S1 for summary of all possible sequence configurations per condition).

familiarized grammar but composed of novel element variants [generalization sequences (“GS”)], and (iii) sequences in which the variants were novel but the combination of categories was inconsistent with the familiar grammar [violation sequences (“VS”)]. We predicted that if individuals processed the grammar, then sequences that deviate from those they are familiar with would elicit a stronger behavioral response [representing a “violation of expectation” (25, 26)], i.e., the amount of time spent looking toward the speaker should follow the pattern of VS > GS ≥ FS. Human subjects participated in an explicit version of the same task using only a Non-AD grammar to ensure that these grammars are, at the very least, readily learnable by linguistically capable individuals.

RESULTS

We found evidence that each species (marmosets: $N = 16$, AD condition $N = 8$, Non-AD condition $N = 8$; chimpanzees: $N = 17$, AD condition $N = 9$, Non-AD condition $N = 8$; humans: $N = 24$, Non-AD only) was able to both generalize the grammars to novel acoustic stimuli and detect when the grammars had been violated, indicating that they processed the structural relationship between dependent elements. In line with our predictions, in each condition, chimpanzees and marmosets spent, on average, twice as long looking toward the speaker after VS than FS or GS (Fig. 2), while reactions to FS and GS did not reliably differ from one another. Among a suite of Bayesian mixed-effects models, the best fitting model for each animal species and each condition (AD and Non-AD), as determined by Watanabe-Akaike information criterion (WAIC) weight [representing the relative likelihood that a model is the best fit in a given set (27)], was

one in which FS and GS did not systematically vary in their effect on gaze duration, whereas VS varied from both (Table 1). In each case, the estimated posterior distribution of gaze duration for VS indicated that this combination of stimuli elicited a longer average looking response than those that were consistent with the familiar grammar (Table 2). The second best-fitting model in each condition was one in which all stimulus types elicited different responses, the outputs of which were consistent with the corresponding best-fitting model (table S2), suggesting that regardless of model uncertainty, the overall pattern of results is robust.

While both nonhuman primate species demonstrated similar patterns of behavioral response by reacting most strongly to VS relative to GS and FS in both AD and Non-AD conditions, marmosets tended to give shorter responses, relative to chimpanzees, for all stimuli in all conditions (Fig. 2), which may reflect between-species differences in sensitivity to artificial auditory stimuli. To explore this further, we pooled the data from each species and included an interaction term between species and stimulus type in the previous best-fitting model structure for each condition. However, for both AD and Non-AD conditions, there was no interaction between species and stimulus type [AD: Beta, 0.75; 89% credible interval (CI), -0.94 to 2.44; Non-AD: Beta, 1.03; 89% CI, -0.91 to 2.81], indicating that the strength of the effect elicited by VS did not substantially differ between species.

A binomial mixed-effects model also found that human participants tested on an explicit version of the Non-AD condition were able to categorize whether a sequence was consistent with those heard in the familiarization period or violated that pattern at well above the 50% chance level (posterior estimated mean accuracy: 82.29%; 89% CI: 75.58 to 88.59%). This confirms that our human

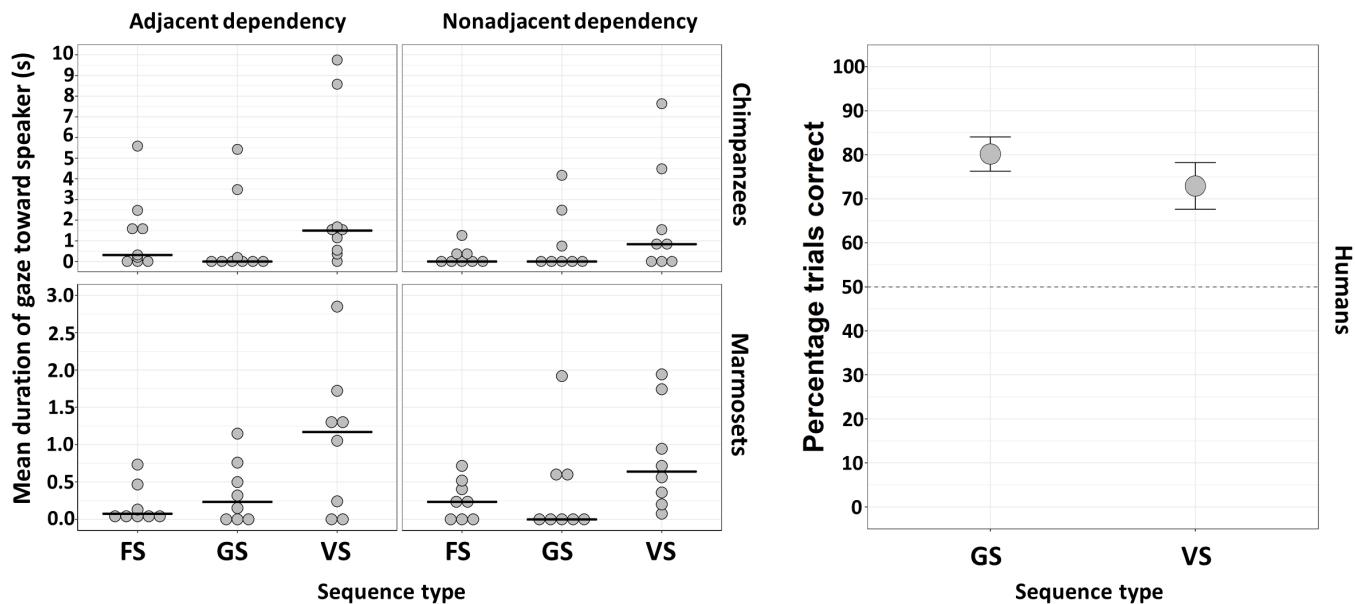


Fig. 2. Participant responses for each experiment. Left: Mean duration spent looking toward the speaker for each primate species in each condition and each sequence type: dots = individuals; bars = group. Right: Response accuracy for human subjects: Bars represent SE.

participants, like marmosets and chimpanzees, were able to abstract the rules governing the composition of these sequences and recognize when they had been violated.

DISCUSSION

In language, monitoring syntactic dependencies between words is cognitively demanding (28) yet central to both its acquisition and processing. Unpacking the origins of this computational capacity is therefore key to a holistic understanding of language and its evolution. We found notable similarities in the ability of marmosets, chimpanzees, and humans to process Non-ADs, providing crucial insights into the evolutionary emergence of this key faculty underpinning human language. Specifically, we present the first evidence of auditory Non-AD processing in chimpanzees, which confirms, in conjunction with our directly comparable marmoset data, that this capacity did not evolve convergently in humans and non-human primates but rather has ancestral origins dating back at least ~40 million years.

There was a large amount of individual variation in each species, with some marmosets giving relatively weak reactions to all sequence types and some chimpanzees not reacting at all (4 of 17 individuals; see fig. S1). Human performance was also subject to individual differences, with 20% of participants failing to perform at above chance level. Whether, in nonhuman animals, this variation was due to subjects not being able to process the structural patterns or simply lacking sufficient motivation to look toward the speaker when distracted by other behaviors is unclear in a passive-response paradigm. However, in previous artificial grammar studies with both humans (17, 29–31) and nonhuman animals (32), there was also substantial individual variation in the ability to learn the underlying rules, suggesting that individual differences in aptitude for auditory pattern recognition may exist across species. Integrating measures of motivation and working memory (31) in future work could begin to elucidate the factors driving such within-species variation. It is also worth noting

that, while in our passive-listening design it was necessary to keep the number of test trials low to maximize the novelty of all test stimuli, the individual differences we identified may be less pronounced in active-task paradigms, which make use of a more extensive number of trials [e.g., Jiang *et al.* (9)] and thereby be less susceptible to noise in the data. It is also possible that our paradigm was confounded by the fact that some animal individuals, like humans, are less sensitive to pitch variation than their conspecifics (33). Paradigms in which subjects must generalize across non-pitch-related acoustic features would therefore be useful in both eliminating this confound and exploring the ways in which nonhumans are capable of abstracting acoustic relationships. Despite the individual differences we observed, our data demonstrate that processing Non-ADs is, minimally, well within reach of at least some individuals in each species without explicit training.

While previous studies have provided evidence for Non-AD processing in nonhuman primates and indeed other animals [whether specifically tested for (5, 7, 8, 13, 15) or when exploring more complex computational capabilities (9, 13, 14)], we propose the extent to which these experiments can reliably inform the evolutionary roots of dependency processing has been complicated by two key issues, which our design circumvents. First, the format of previously used grammars often could not conclusively rule out more parsimonious alternative explanations. For example, dependencies implemented in some grammars were between two near-identical elements (5, 6), which could be parsed using relatively simple heuristics such as “Sound A must occur more than once” rather than processing a positional relationship between the two elements (34). Another possibility is that acoustic similarities between test and training sequences allowed subjects to generalize based on this perceptual feature (4, 34). We actively excluded such alternative explanations by using a paradigm in which (i) the dependencies were between different categories of sounds rather than repetitions of the same sound and (ii) the test items were different from the learning items. Furthermore, these design features meant that individuals were exposed to

Table 1. Relative fits (in descending order of WAIC weight) for each statistical model corresponding to our a priori hypotheses regarding patterns of gaze duration (see the “Statistical analysis” section in Materials and Methods). “Varies” does not assume directionality in relation to other sequence types; this was instead determined by inspecting posterior distributions.

Model	Species	Condition	WAIC	WAIC weight
FS = GS, VS varies	Marmosets	Nonadjacent	138.5	0.67
FS, GS, and VS vary			140.5	0.24
FS = VS, GS varies			143.2	0.06
GS = VS, FS varies			146.3	0.01
Null model (FS = GS = VS)			147.1	0.01
FS = GS, VS varies	Marmosets	Adjacent	189.6	0.52
FS, GS, and VS vary			191.0	0.25
FS = VS, GS varies			191.5	0.20
GS = VS, FS varies			196.9	0.01
Null model (FS = GS = VS)			196.9	0.01
FS = GS, VS varies	Chimpanzees	Non-adjacent	239.1	0.91
FS, GS, and VS vary			243.9	0.09
Null model (FS = GS = VS)			255.8	0.00
FS = VS, GS varies			264.9	0.00
GS = VS, FS varies			268.4	0.00
FS = GS, VS varies	Chimpanzees	Adjacent	332.8	0.82
FS, GS, and VS vary			337.7	0.07
GS = VS, FS varies			338.1	0.06
Null model (FS = GS = VS)			339.8	0.03
FS = VS, GS varies			340.9	0.01

Table 2. Estimated posterior distributions for best-fitting model of each condition presented in Table 1. Coefficients represent predicted total time spent looking toward the speaker after FS/GS, and VS relative to FS/GS.

Species	Condition	Parameter	Mean	SD	5.5% CI	94.5% CI
Marmosets	Nonadjacent	FS/GS	0.29	0.15	0.07	0.53
		VS	0.61	0.28	0.17	1.03
Marmosets	Adjacent	FS/GS	0.30	0.16	0.03	0.52
		VS	0.79	0.38	0.16	1.38
Chimpanzees	Non-adjacent	FS/GS	0.49	0.28	0.01	0.88
		VS	1.54	1.09	-0.19	3.20
Chimpanzees	Adjacent	FS/GS	1.28	0.67	0.20	2.25
		VS	1.27	0.85	0.02	2.60

a large number of sequence configurations during the familiarization phase (64 for each grammar, 128 in total; see Fig. 1). This ultimately minimized the likelihood that, during the test phase, individuals simply reacted to deviations from individually memorized sequences, rather than processing the dependent relationships between first and last sound categories.

Second, there also exists considerable variation in the range of methodologies and paradigms implemented across previous Non-AD processing studies in nonhuman animals. For example, a number of paradigms used passive learning of grammars [e.g., cotton-top

tamarins, *Saguinus oedipus* (8, 24); squirrel monkeys, *Saimiri sciureus* (5)], while others actively trained subjects via operant conditioning [e.g., starlings, *Sturnus vulgaris* (35); rhesus macaques, *Macaca mulatta* (9)]. In addition, some studies investigated sequence processing in the visual [e.g., chimpanzees (15, 16); rats, *Rattus norvegicus* (36); cotton-top tamarins (12)] rather than auditory domain [e.g., Bengalese finches, *Lonchura striata domestica* (14)], and certain studies constructed their grammars from artificial stimuli [e.g., squirrel monkeys and common marmosets (5, 6)] as opposed to using vocalizations from the study species’ own repertoire [e.g., starlings (13, 35)].

While experimental design must account for the pertinent cognitive, behavioral, and morphological profiles of a study species, it is advantageous to keep all other factors as similar as possible when aiming to facilitate fair comparisons across species (19–23). Not doing so may confound a directly comparative approach, which is central to unpacking the evolutionary roots of Non-AD processing and the corresponding selective pressures. To this end, we designed a standardized paradigm using identical stimuli, grammars, protocol, and response measures that is nevertheless flexible enough to be applied across species, thus ensuring that any similarities we identified between apes and monkeys in their behavioral responses were likely to reflect shared cognitive mechanisms for processing Non-ADs.

Of course, the Non-ADs implemented in our artificial grammars are simplistic in structure compared to some of the complex dependencies frequently produced and processed in human language (2). In the future, it would therefore be worthwhile to extend this paradigm to include features such as variable distances between dependent elements (6) and multiple layers of embedding (37, 38) or explore the influence of features that promote or inhibit artificial grammar learning [e.g., “edge effects” in dependency positioning (39)]. Previous work has also shown that, in humans, the semantic layer of language makes complex grammars considerably easier to learn (40). Incorporating this factor into animal paradigms may therefore be fruitful in exploring the additive influence of meaning in sequence processing capabilities, although doing so is not without its own challenges (41). Together, such work may help tease apart between-species differences in the computational limits of Non-AD processing, allowing a yet more detailed picture of the extent of phylogenetic continuities in this capacity.

Previous work has suggested that Non-AD processing in language may be a novel application of an evolutionarily ancient ability for computing complex spatial, temporal, or social relationships/hierarchies (15, 26, 42, 43). Our principal finding that the evolutionary origins of Non-AD processing predates the evolution of language provides key support for this hypothesis. However, while the basic capacity for Non-AD processing appears to be widespread, an open question remains regarding whether humans are indeed unique in the capacity to produce Non-ADs in a communicative context. Previous work suggests that both chimpanzees and marmosets combine calls from their natural vocal repertoire into larger structures (44, 45), although currently there is only evidence for the existence of simple combinations (two call types, i.e., bigrams). Follow-up work combining standardized artificial grammar experiments, such as those proposed here, with detailed analysis of statistical relationships between adjacent and nonadjacent calls in animal vocal sequences will help shed light on this issue.

MATERIALS AND METHODS

Subject details

Experiment 1—Humans

In total, 24 individuals (14 female and 10 male) with a mean age of 25.6 years ($SD = 4.47$) participated in the experiment. All participants had at least some competency in a second language. All participants had normal hearing and no history of neurological disorders. Experiment 1 was carried out at the University of Osnabrück, Germany. Our study was approved by the ethics committee of the University of Osnabrück and conformed to the guidelines of the Declaration of Helsinki (2013).

Experiment 2—Common marmosets (2a) and chimpanzees (2b)

Experiment 2a. A total of 16 adult common marmosets participated in the study (AD condition: $N = 8$, 4 female; Non-AD condition: $N = 8$, 5 female). Full demographic information can be found in table S3. All individuals were housed at the Primate Station of the University of Zurich. Animals in the AD condition were housed in pairs in enclosures composed of an indoor area of 1 m (width) by 2 m (depth) by 2 m (height) and an outdoor area of 1.8 m by 2.4 m by 3.2 m. Marmosets from the Non-AD condition were housed in family groups in enclosures composed of an indoor area of 1.8 m by 2.4 m by 2.7 m and an outdoor area of 1.8 m by 2.4 m by 3.2 m. All researchers were compliant with the Swiss regulations regarding ethical treatment of animals in experiments, and the experiments were approved by the Kantonale Veterinäramt Zürich, license number ZH223/16.

Experiment 2b. Subjects were 17 adult chimpanzees (AD condition: $N = 9$, 4 female; Non-AD condition: $N = 8$, 5 female) housed at the National Center for Chimpanzee Care located at the Michale E. Keeling Center for Comparative Medicine and Research of The University of Texas MD Anderson Cancer Center (UTMDACC) in Bastrop, Texas. Ethical approval for this study was granted by the Institutional Animal Care and Use Committee of the UTMDACC, adhering to all the legal requirements of the U.S. law and the American Society of Primatologists’ principles for the ethical treatment of non-human primates. Individuals were housed in mixed-sex groups of between four and eight individuals with access to both indoor (two “dens” of 14 m^2 each) and outdoor (90 or 400 m^2 , depending on group size) enclosures. All subjects voluntarily participated in the testing procedures. Full demographic information can be found in table S4.

Methods details

Stimuli

Each of our artificial grammar sequences was composed of elements drawn from six computer-generated acoustic categories (A, B, C, D, X1, and X2). AD grammars were composed of one element from each of two dependent categories (A, B, C, or D), and Non-AD grammars were identical but with the addition of a central “X” element (from category X1 or X2), separating the two dependent elements (Fig. 1). For each condition (AD and Non-AD), we constructed two sets of paired grammars (table S1). In one set (Grammar 1), A elements were always followed by B elements (Grammar 1a), and C elements were predictive of D (Grammar 1b). In the second set, the roles of B and D elements were reversed (Grammar 2), with C dependent on A (Grammar 2a) and B dependent on C (Grammar 2b). To control for the possibility that certain sound pairings might be relatively easier to learn, assignment to Grammar 1 and Grammar 2 was counterbalanced across participants as much as possible, given the group sizes available, within conditions for all species (tables S3 and S4). Each acoustic category, including both X1 and X2, was composed of 16 pitch-shifted variants, half of which were used to construct familiarization sequences (“FS”) and half for GS and VS so that these were novel to the listener (Fig. 1 and table S1). Each category variant was separated by 50 Hz, starting from 500 Hz (at onset). An exception was the frequency difference between variants 8 and 9 of each category, which were separated by a 200-Hz gap to increase the perceptual difference between ranges 1 to 8 and 9 to 16. Osmanski *et al.* (46) have demonstrated that common marmosets are reliably sensitive to pitch differences of at least 42 Hz (at 220 Hz), becoming more sensitive as pitch increases. Because the minimum pitch of our stimuli was 500 Hz and increased in 50 Hz steps, the differences between

tones would have been discriminable to this species. Crucially, the jump between FS and GS (a 200-Hz difference between the highest familiarization tone and the lowest generalization tone) would have been highly salient. Chimpanzees and humans are known to have even more sensitive pitch discrimination than marmosets, particularly at these lower ranges (47). Within a sequence type, these variants were not restricted in how they could be combined. For example, for an FS of the AD condition, A-1 could be followed by any of B-1 to B-8 (see Fig. 1 examples). Because “local redundancy” (a measure of nongrammatical predictability between elements) (48) has been proposed as a confounding variable explaining apparent processing of artificial grammars, we used the methods of Jamieson *et al.* (48) to examine whether this applied to our own design. We found that our sequences had an extremely low measure of local redundancy (0.05), which, crucially, did not differ between FS, GS, and VS. Examples of the scripts used to generate these acoustic elements in Praat are available at <https://osf.io/4m3gv/>.

All elements had a duration of 1500 ms, with a 10-ms volume fade in/out to eliminate sound onset effects. All elements were generated using Praat (49). For all sequences in the human experiment, there was a 250-ms gap between elements. For the marmoset and chimpanzee experiments, this was lengthened to 500 ms to increase the saliency of the individual elements and eliminate overlap from echoes caused by the acoustic environments in which the animals were housed. Therefore, including gaps between sounds, all AD sequences lasted a total of 3500 ms, and Non-AD sequences lasted 5500 ms (5000 ms for humans). The full lists of sequences used for familiarization and test phases can be found at our online repository: <https://osf.io/4m3gv/>.

Experiment 1 protocol

While the human capacity to process Non-ADs is not in doubt, they were included in this study primarily to validate our artificial grammars before being tested on animal subjects. Although an identical passive response paradigm would have been advantageous, implementing this with adult human subjects was not feasible because of, for example, the extensive habituation phase necessary for such a paradigm, and an explicit choice task was used instead. Participants were seated in front of the experimental computer and provided with written instructions outlining the task in either English or German. When they were ready to begin, participants were exposed to four experimental blocks, each of which was composed of a familiarization phase and a test phase. In each familiarization phase, participants were played 60 FS (30 of each grammar). The participants had been instructed to listen carefully to these sequences and try to identify any rules that the sequences followed. Upon completion of the familiarization phase, a message notified participants that they were about to begin the test phase. During a test phase, participants were played a pseudo-randomized (such that no more than three of a sequence type occurred in a row) list of 12 GS and 12 VS. After each sequence, an on-screen prompt appeared asking the participant to provide a “Yes” or “No” response via button press as to whether they thought the test sequence followed the rules they had identified during the familiarization phase. Upon completion of the test phase, a new familiarization phase began until four experimental blocks had been completed.

Experiment 2a and 2b protocol

Familiarization phase

All animal subjects were provided with at least 10 familiarization sessions, carried out twice per day for 5 days. In each familiarization

session, we played 240 FS (120 of each grammar) composed of pseudo-randomized variants of the acoustic elements in a group context. Lists were constrained to have an equal number of each variant type and for no more than four sequences of the same grammar to play in a row. There was a 2500-ms gap between sequences. When possible, one session was carried out in the morning, and the second was carried out in the afternoon. In cases where circumstances prevented this, both sessions were carried out in the morning or afternoon but separated by at least 1 hour. For the AD conditions, a familiarization session took 24 min to play from start to finish. In the Non-AD conditions, because of the additional acoustic element in each sequence, the total duration was 32 min. The position of the speaker was alternated between left and right sides of the enclosure in consecutive sessions.

Species-specific details: Marmosets

All familiarization and test sessions for marmosets were carried out in an enclosure separated from their home enclosure, sized 1 m by 2 m by 2 m. This was a practical necessity for the speaker setup and also to ensure that a subject’s groupmates were acoustically isolated during test sessions. Familiarization sessions were carried out in a group context. During test sessions, individuals voluntarily entered a smaller compartment with dimensions of 40 cm by 40 cm by 75 cm, the bottom of which was placed approximately 1.5 m from ground level. The compartment had a wooden board covering the bottom, which was covered in approximately 3 to 4 cm of mulch for comfort, above which a hanging perch was also placed. The purpose of this smaller compartment was to ensure that the focal individual was within camera frame and in line of sight to the speaker at all times during the test phase.

Species-specific details: Chimpanzees

All familiarization and test sessions were carried out in a group’s indoor enclosure. Volume levels were controlled such that the sequences were only clearly audible when an individual was inside or standing in the doorway to the outdoor enclosure. Individuals were free to move between indoor and outdoor enclosures at all times.

Test phase

Immediately before test sessions, subjects were exposed to 60 familiarization trials (30 of each grammar) to refamiliarize them with the grammars. After this “refamiliarization phase,” the experimenter waited at least 2 min before beginning the test phase. Test sessions were composed of 12 trials in total, with 4 of each stimulus type (4 FS, 4 GS, and 4 VS). This number of trials was chosen to minimize the risk that subjects habituated to the GS and VS sequences within the experiment itself while also remaining consistent with similar paradigms (6, 24). Half of these trials corresponded to each grammar (i.e., AXB/CXD). Trials were marked as invalid if the subject left the camera frame, an external noise distracted the subject, or they were looking in the direction of the speaker at the onset of the final sound in a sequence. Subjects were only included for analysis if they had at least one valid trial for each stimulus type. The onset of test trials was activated using a remote control concealed behind the experimenter’s back. Lists of test sequences were pseudo-randomized so that no more than three trials of the same type were carried out in succession. Sequences presented in the test phase were balanced in frequency distance between their constituent elements. In other words, if an FS in the AD test phase was composed of A₁-B₈ (where 1 and 8 refer to pitch variants), then there existed a corresponding GS of A₉-B₁₆ and a VS of A₉-D₁₆. The position of GS and VS within each list was counterbalanced across subjects. The position of the speaker (left or right of enclosure) was also counterbalanced across subjects during test phases. During all test trials, the experimenter

stood directly behind the camera and looked toward the floor so as to avoid providing cues to the subject. Each test trial would be triggered only when the subject was (i) stationary, (ii) within frame of the camera, (iii) facing either the camera or at least 90° away from the speaker, and (iv) at least 15 s had passed since the last trial. The volume was set at a level such that the stimuli were inaudible to the rest of the subject's group. Clips from violation trials in the Non-AD condition for a marmoset and a chimpanzee can be found at our online repository: <https://osf.io/4m3gv/>.

Test phase species-specific details: Marmosets

During test sessions, marmosets were provided with either mealworms or tree gum paste as a distractor activity. This was placed in the compartment to encourage foraging and avoid subjects attending to the experimenter. The food type was dependent on the preferences of the individual (if they did not like mealworms, then sap was provided) and is indicated in table S3. Two individuals ("NAND" and "GAR") frequently produced contact calls upon being separated from their group, so a second individual (who had already been tested in the same condition) was brought into an adjacent enclosure during testing, on the opposite side to the speaker, to minimize arousal. The experimenter only started the onset of test trials when the second individual was silent and stationary.

Test phase species-specific details: Chimpanzees

During the test phase, chimpanzees were encouraged to move into their indoor area in pairs (to avoid stress resulting from social isolation), and the door was closed behind them to prevent interruption from groupmates. Where possible, both individuals were naïve to the test stimuli. However, after an individual had been tested, they were still eligible to be partnered with other naïve individuals in the group. Preliminary work indicated that because of active moving, obtaining obvious looking responses during playbacks was logically complicated. To circumvent this during test sessions, each chimpanzee was provided with a bottle of diluted juice attached to the mesh of their enclosure at a height easily reached by a seated chimpanzee. This ensured that subjects remained in-frame of the camera and encouraged them to orient their heads directly forward so that they were reliably looking away from the speaker before a trial was triggered. Because of the layout of the enclosure, it was not possible to position the speaker directly 90° from where the subjects were situated, so instead, it was placed as close as possible to this at a distance approximately 3 to 4 m from the subjects so that a distinct movement of the head was still required to look at it.

Apparatus

Experiment 1. The experiment was coded in MATLAB R2017B version 9.3.0.713579 with the Psychophysics Toolbox add-on (50) and run through a desktop computer. Audio stimuli were played in stereo through two speakers positioned on either side of the monitor. Participants inputted their responses through two stand-alone buttons placed on the table in front of them. Participants were seated between 65 and 70 cm from the speakers and monitor in front of them.

Experiments 2a and 2b. All stimuli were played through a Braven BRV-X "ultra-rugged" series speaker using a slightly adapted version of the MATLAB script from Experiment 1 on a Windows laptop. All familiarization and test sessions were recorded with Sony HDR-CX240E digital video cameras.

Quantification and statistical analysis

Video coding

All videos were coded frame by frame using BORIS behavioral observation software (v 6.2.2 available at www.boris.unito.it/) by SKW. After

the onset of the final element in each test sequence, the coder placed a time stamp on frames in which the subject oriented their head directly toward the speaker (see fig. S2) and a second time stamp when their head was no longer oriented toward the speaker. The duration between these two time stamps was then calculated. The total additive duration of each looking bout occurring within 15 s of the onset of the final element in a test sequence was used as our response measure. We selected this response window length given that subjects were engaged in foraging behavior during testing, and because the stimuli did not have any ecological urgency for them, we did not necessarily expect an immediate behavioral response. A 15-s response window was therefore judged to plausibly minimize the number of false negatives, while false positives were already controlled for by the relatively low likelihood that subjects would look precisely toward the speaker if picking a location at random to focus their gaze.

For marmosets, an unambiguous look-to-speaker was coded as when an individual's head was oriented fully and directly toward the speaker. For chimpanzees, because of the position of the speaker and the fact they were engaged in sucking juice from a straw, a change in head orientation of 45° to 90° toward the speaker was coded as a look. Interobserver reliability tests were carried out on 50% of all trials. To reduce bias, the independent observer was provided with muted videos to ensure that they were blind to both condition and stimulus type. Correlation analyses suggested strong agreement between observers ($r = 0.82$ for marmosets and 0.92 for chimpanzees).

Statistical analysis

For Experiment 1, we used Bayesian binomial mixed-effects models to estimate the posterior distribution of response accuracies for (i) all stimuli, (ii) only generalization stimuli, and (iii) only violation stimuli. Random intercepts were fitted for individual identity and test block to control for the fact that multiple data points were drawn from each level of each of these clusters.

For Experiment 2, we used Bayesian Markov Chain Monte Carlo generalized linear mixed-effects models with a Gaussian distribution to examine the differences in time spent looking toward the speaker in the 15-s response window following each test trial (Familiar, Generalization, Violation). All models were implemented in R with the "map2stan" function from the "Rethinking" package (51). We took a model comparison approach, which entailed fitting five different models specified according to the predictions of a priori hypotheses regarding the possible outcomes of the experiments. These predictions were as follows:

- 1) That FS, GS, and VS did not systematically differ in their effect on looking behavior.
- 2) That FS, GS, and VS all varied independently in their effect on looking behavior.
- 3) That FS and GS elicited a similar response, but VS differed from this.
- 4) That FS and VS elicited a similar response, but GS differed from this.
- 5) That GS and VS elicited a similar response, but FS differed from this.

To demonstrate that a given species processes a dependency (AD or Non-AD), we reasoned that predictions 2 or 3 would need to be supported, with VS eliciting a stronger looking response than the other sequence types, as determined by inspection of the respective posterior distributions and 89% CIs (52). These five models were compared using WAIC (52) scores to determine which provided the best model fit for each species and condition. We fitted random intercepts

and slopes for identity to allow individual differences and the fact that multiple data points were drawn from each individual.

All models were run with two chains of 20,000 iterations and a warm-up period of 1000 iterations. Trace plots, rhat values, and effective sample sizes were used to assess model convergence. We fitted the models using weakly informative (or regularizing) priors to mitigate potential overfitting.

All analyses were carried out using R (53) and RStudio (54) with packages “rstan” (55) and “rethinking” (51). All figures were drawn using “ggplot2” (56).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/43/eabb0725/DC1>

REFERENCES AND NOTES

1. M. D. Hauser, N. Chomsky, W. T. Fitch, The faculty of language: What is it, who has it, and how did it evolve? *Science* **298**, 1569–1579 (2002).
2. B. Wilson, M. Spierings, A. Ravignani, J. L. Mueller, T. H. Mintz, F. Wijnen, A. van der Kant, K. Smith, A. Rey, Non-adjacent dependency learning in humans and other animals. *Top. Cogn. Sci.* **12**, 843–858 (2018).
3. A. S. Reber, Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* **6**, 855–863 (1967).
4. G. J. L. Beckers, R. C. Berwick, K. Okanoya, J. J. Bolhuis, What do animals learn in artificial grammar studies? *Neurosci. Biobehav. Rev.* **81**, 238–246 (2017).
5. A. Ravignani, R.-S. Sonnweber, N. Stobbe, W. T. Fitch, Action at a distance: Dependency sensitivity in a New World primate. *Biol. Lett.* **9**, 20130852 (2013).
6. S. A. Reber, V. Šlipogor, J. Oh, A. Ravignani, M. Hoeschele, T. Bugnyar, W. T. Fitch, Common marmosets are sensitive to simple dependencies at variable distances in an artificial grammar. *Evol. Hum. Behav.* **40**, 214–221 (2019).
7. A. E. Milne, J. L. Mueller, C. Männel, A. Attaheri, A. D. Friederici, C. I. Petkov, Evolutionary origins of non-adjacent sequence processing in primate brain potentials. *Sci. Rep.* **6**, 36259 (2016).
8. E. L. Newport, M. D. Hauser, G. Spaepen, R. N. Aslin, Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognit. Psychol.* **49**, 85–117 (2004).
9. X. Jiang, T. Long, W. Cao, J. Li, S. Dehaene, L. Wang, Production of supra-regular spatial sequences by macaque monkeys. *Curr. Biol.* **28**, 1851–1859.e4 (2018).
10. R. Malassis, A. Rey, J. Fagot, Non-adjacent dependencies processing in human and non-human primates. *Cognit. Sci.* **42**, 1677–1699 (2018).
11. A. Rey, P. Perruchet, J. Fagot, Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (*Papio Papio*). *Cognition* **123**, 180–184 (2012).
12. E. Versace, J. R. Rogge, N. Shelton-May, A. Ravignani, Positional encoding in cotton-top tamarins (*Saguinus oedipus*). *Anim. Cogn.* **22**, 825–838 (2019).
13. J. Chen, C. ten Cate, Bridging the gap: Learning of acoustic nonadjacent dependencies by a songbird. *J. Exp. Psychol. Anim. Learn. Cogn.* **43**, 295–302 (2017).
14. K. Abe, D. Watanabe, Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* **14**, 1067–1074 (2011).
15. R. Sonnweber, A. Ravignani, W. T. Fitch, Non-adjacent visual dependency learning in chimpanzees. *Anim. Cogn.* **18**, 733–745 (2015).
16. A. Ravignani, R. Sonnweber, Chimpanzees process structural isomorphisms across sensory modalities. *Cognition* **161**, 74–79 (2017).
17. J. L. Mueller, A. D. Friederici, C. Männel, Auditory perception at the root of language learning. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15953–15958 (2012).
18. R. Frost, B. C. Armstrong, N. Siegelman, M. H. Christiansen, Domain generality versus modality specificity: The paradox of statistical learning. *Trends Cogn. Sci.* **19**, 117–125 (2015).
19. L. G. Dean, R. L. Kendal, S. J. Schapiro, B. Thierry, K. N. Laland, Identification of the social and cognitive processes underlying human cumulative culture. *Science* **335**, 1114–1118 (2012).
20. J. M. Burkart, C. van Schaik, Group service in macaques (*Macaca fuscata*), capuchins (*Cebus apella*) and marmosets (*Callithrix jacchus*): A comparative approach to identifying proactive prosocial motivations. *J. Comp. Psychol.* **127**, 212–225 (2013).
21. ManyPrimates, D. M. Altschul, M. J. Beran, M. Bohn, J. Call, S. DeTroy, S. J. Duguid, C. L. Egelkamp, C. Fichtel, J. Fischer, M. Flessert, D. Hanus, D. B. M. Haun, L. M. Haux, R. A. Hernandez-Aguilar, E. Herrmann, L. M. Hopper, M. Joly, F. Kano, S. Keupp, A. P. Melis, A. M. Rodrigo, S. R. Ross, A. Sánchez-Amaro, Y. Sato, V. Schmitt, M. K. Schweinfurth, A. M. Seed, D. Taylor, C. J. Völter, E. Warren, J. Watzek, Establishing an infrastructure for collaboration in primate cognition research. *PLOS ONE* **14**, e0223675 (2019).
22. F. Kano, C. Krupenye, S. Hirata, M. Tomonaga, J. Call, Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20904–20909 (2019).
23. C. Krupenye, F. Kano, S. Hirata, J. Call, M. Tomasello, Great apes anticipate that other individuals will act according to false beliefs. *Science* **354**, 110–114 (2016).
24. W. T. Fitch, M. D. Hauser, Computational constraints on syntactic processing in a nonhuman primate. *Science* **303**, 377–380 (2004).
25. E. S. Spelke, K. Breinlinger, J. Macomber, K. Jacobson, Origins of knowledge. *Psychol. Rev.* **99**, 605–632 (1992).
26. R. M. Seyfarth, D. L. Cheney, T. J. Bergman, Primate social cognition and the origins of language. *Trends Cogn. Sci.* **9**, 264–266 (2005).
27. M. R. E. Symonds, A. Moussalli, A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behav. Ecol. Sociobiol.* **65**, 13–21 (2011).
28. E. Bach, C. Brown, W. Marslen-wilson, Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Lang. Cognit. Process.* **1**, 249–262 (1986).
29. A. Ravignani, P. Filippi, W. T. Fitch, Perceptual tuning influences rule generalization: Testing humans with monkey-tailored stimuli. *J. Perception* **10**, 2041669519846135 (2019).
30. N. Siegelman, R. Frost, Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *J. Mem. Lang.* **81**, 105–120 (2015).
31. J. B. Misayak, M. H. Christiansen, Statistical learning and language: An individual differences study. *Lang. Learn.* **62**, 302–331 (2012).
32. C. A. van Heijningen, J. de Visser, W. Zuidema, C. ten Cate, Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20538–20543 (2009).
33. J. R. Engle, S. Tinling, G. H. Recanzone, Age-related hearing loss in rhesus monkeys is correlated with cochlear histopathologies. *PLOS ONE* **8**, e55092 (2013).
34. S. Ghirlanda, Can squirrel monkeys learn an ABⁿA grammar? A re-evaluation of Ravignani et al. (2013). *PeerJ* **5**, e3806 (2017).
35. T. Q. Gentner, K. M. Fenn, D. Margoliash, H. C. Nusbaum, Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204–1207 (2006).
36. R. A. Murphy, E. Mondragón, V. A. Murphy, Rule learning by rats. *Science* **319**, 1849–1851 (2008).
37. M. Winkler, J. L. Mueller, A. D. Friederici, C. Männel, Infant cognition includes the potentially human-unique ability to encode embedding. *Sci. Adv.* **4**, eaar8334 (2018).
38. A. D. Friederici, J. L. Mueller, R. Oberecker, Precursors to natural grammar learning: Preliminary evidence from 4-month-old infants. *PLOS ONE* **6**, e17920 (2011).
39. J. L. Mueller, J. Bahlmann, A. D. Friederici, Learnability of embedded syntactic structures depends on prosodic cues. *Cognit. Sci.* **34**, 338–349 (2010).
40. A. Fedor, M. Varga, E. Szathmáry, Semantics boosts syntax in artificial grammar learning tasks with recursion. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 776–782 (2012).
41. S. A. Reber, M. Boeckle, G. Szuml, J. Janisch, T. Bugnyar, W. T. Fitch, Territorial raven pairs are sensitive to structural changes in simulated acoustic displays of conspecifics. *Anim. Behav.* **116**, 153–162 (2016).
42. K. Zuberbühler, Evolutionary roads to syntax. *Anim. Behav.* **151**, 259–265 (2019).
43. P. Katerina, A. Yiannis, The minimalist grammar of action. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 103–117 (2012).
44. J. A. Agamaite, C.-J. Chang, M. S. Osmanski, X. Wang, A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*). *J. Acoust. Soc. Am.* **138**, 2906–2928 (2015).
45. C. Crockford, C. Boesch, Call combinations in wild chimpanzees. *Behaviour* **142**, 397–421 (2005).
46. M. S. Osmanski, X. Song, Y. Guo, X. Wang, Frequency discrimination in the common marmoset (*Callithrix jacchus*). *Hear. Res.* **341**, 1–8 (2016).
47. S. Kojima, Comparison of auditory functions in the chimpanzee and human. *Folia Primatol.* **55**, 62–72 (1990).
48. R. K. Jamieson, U. Nevzorova, G. Lee, D. J. K. Mewhort, Information theory and artificial grammar learning: Inferring grammaticality from redundancy. *Psychol. Res.* **80**, 195–211 (2016).
49. P. Boersma, D. Weenink, *Praat: Doing Phonetics by Computer* (Version 5.1.13) (2009); www.praat.org.
50. D. H. Brainard, The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
51. R. McElreath, *Rethinking: Statistical Rethinking Book Package* (2019).
52. R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (CRC Press, 2016), vol. 122.
53. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2018); www.R-project.org/.
54. RStudio Team, *RStudio: Integrated Development for R* (2015); www.rstudio.com.

55. J. Guo, J. Gabry, B. Goodrich, *rstan: R Interface to Stan* (2018); <https://CRAN.R-project.org/package=rstan>.
56. H. Wickham, W. Chang, L. Henry, T. L. Pedersen, K. Takahashi, C. Wilke, K. Woo, *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (2018); <https://CRAN.R-project.org/package=ggplot2>.

Acknowledgments: We thank S. Sehner and A. Bosshard for carrying out interobserver reliability testing. We thank G. Bazzell at the UZH primate station and all the primate care staff at MD Anderson Cancer Center who facilitated this research and provided their expertise whenever required. We also thank S.-Y. Chong and M. Maschelloni for their assistance in designing, programming, and carrying out the human experiment; B. Bickel, S. Sauppe, C. van Schaik, K. Slocombe, D. Blasi, S. Stoll, Z. Oldfield, P. Filippi, and S. Engesser for discussions; and two anonymous reviewers for their constructive feedback on previous drafts of this manuscript.

Funding: S.K.W., S.W.T., and J.M.B. were funded by the Swiss National Science Foundation (S.K.W. and S.W.T.: grant PP00P3_163850; J.M.B.: grant 31003A_172979). S.K.W. and S.W.T. were also funded by NCCR Evolving Language, Swiss National Science Foundation Agreement #51NF40_180888. The chimpanzees at MD Anderson Cancer Center were funded by

cooperative agreement U42 OD-011197. **Author contributions:** Data curation: S.K.W.; formal analysis: S.K.W.; investigation: S.K.W.; methodology: S.K.W., J.M.B., J.L.M., and S.W.T.; project administration: S.K.W., J.M.B., S.J.S., S.P.L., and S.W.T.; software: S.K.W. and J.L.M.; visualization: S.K.W.; writing: S.K.W., J.M.B., S.J.S., S.P.L., J.L.M., and S.W.T.; resources: J.M.B., S.J.S., S.P.L., J.L.M., and S.W.T.; conceptualization: S.K.W., J.L.M., and S.W.T.; supervision: S.K.W., J.L.M., and S.W.T.

Competing interests: The authors declare that they have no competing interests. **Data and materials availability:** All raw data and R scripts used for analysis are available to download from the Open Science Framework (URL: <https://osf.io/4m3gv/>).

Submitted 28 January 2020

Accepted 8 September 2020

Published 21 October 2020

10.1126/sciadv.abb0725

Citation: S. K. Watson, J. M. Burkart, S. J. Schapiro, S. P. Lambeth, J. L. Mueller, S. W. Townsend, Nonadjacent dependency processing in monkeys, apes, and humans. *Sci. Adv.* **6**, eabb0725 (2020).

Science Advances

Nonadjacent dependency processing in monkeys, apes, and humans

Stuart K. Watson, Judith M. Burkart, Steven J. Schapiro, Susan P. Lambeth, Jutta L. Mueller and Simon W. Townsend

Sci Adv **6** (43), eabb0725.
DOI: 10.1126/sciadv.eabb0725

ARTICLE TOOLS

<http://advances.science.org/content/6/43/eabb0725>

SUPPLEMENTARY MATERIALS

<http://advances.science.org/content/suppl/2020/10/19/6.43.eabb0725.DC1>

REFERENCES

This article cites 49 articles, 9 of which you can access for free
<http://advances.science.org/content/6/43/eabb0725#BIBL>

PERMISSIONS

<http://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).