

The apes' edge: positional learning in chimpanzees and humans

Ansgar D. Endress · Sarah Carden ·
Elisabetta Versace · Marc D. Hauser

Received: 21 August 2009 / Revised: 26 November 2009 / Accepted: 28 November 2009
© Springer-Verlag 2009

Abstract A wide variety of organisms produce actions and signals in particular temporal sequences, including the motor actions recruited during tool-mediated foraging, the arrangement of notes in the songs of birds, whales and gibbons, and the patterning of words in human speech. To accurately reproduce such events, the elements that comprise such sequences must be memorized. Both memory and artificial language learning studies have revealed at least two mechanisms for memorizing sequences, one tracking co-occurrence statistics among items in sequences (i.e., transitional probabilities) and the other one tracking the *positions* of items in sequences, in particular those of items in sequence-edges. The latter mechanism seems to dominate the encoding of sequences after limited exposure, and to be recruited by a wide array of grammatical phenomena. To assess whether humans differ from other species in their reliance on one mechanism over the other after limited exposure, we presented chimpanzees (*Pan troglodytes*) and human adults with brief exposure to six

items, auditory sequences. Each sequence consisted of three distinct sound types (X, A, B), arranged according to two simple temporal rules: the A item always preceded the B item, and the sequence-edges were always occupied by the X item. In line with previous results with human adults, both species primarily encoded positional information from the sequences; that is, they kept track of the items that occurred in the sequence-edges. In contrast, the sensitivity to co-occurrence statistics was much weaker. Our results suggest that a mechanism to spontaneously encode positional information from sequences is present in both chimpanzees and humans and may represent the default in the absence of training and with brief exposure. As many grammatical regularities exhibit properties of this mechanism, it may be recruited by language and constrain the form that certain grammatical regularities take.

Keywords Serial memory · Artificial grammar learning · Language acquisition

Introduction

Sounds, actions and physical events all unfold over time. Thus, for example, when a songbird or whale sings, it strings together species-specific notes into phrases that, together, provide information on individual, population and species identity (e.g., Prather et al. 2008; Suzuki et al. 2006). Similarly, when a chimpanzee or crow prepares to use a tool for extractive foraging, it must coordinate into a precise sequence the identification of a target resource, the gathering and preparation of a relevant tool, the use of this tool in a particular fashion, often repeating the same action with the tool until the target food is obtained (e.g., Byrne 1999). And when humans perceive speech, they must recall

A. D. Endress (✉)
Departments of Linguistics and Psychology, Harvard University,
33 Kirkland St, Cambridge, MA 02138, USA
e-mail: ansgar.endress@m4x.org

S. Carden · E. Versace
Department of Psychology, Harvard University, 33 Kirkland St,
Cambridge, MA 02138, USA

E. Versace
Department of Psychology, University of Trieste, Trieste, Italy

M. D. Hauser
Departments of Psychology and Human Evolutionary Biology,
Harvard University, 33 Kirkland St, Cambridge,
MA 02138, USA

not only which words were produced, but where in the sequence each word occurred relative to the others, each with a specific meaning. To adaptively generate, perceive and comprehend such events, therefore, animals—humans included—must be equipped with mechanisms that process and memorize sequential input (Conway and Christiansen 2001; Terrace 2005).

Sequences can be memorized in different ways, relying on distinct mechanisms. On the one hand, it is possible to memorize the sequential relations among the items in a sequence. On the other hand, it is possible to encode the *positions* of items in a sequence. In experimental studies of humans, the second encoding mechanism seems to dominate memory processes after limited exposure and seems to be particularly important for language (Endress and Bonatti 2007). Indeed, many grammatical regularities are defined by the positions of certain elements in both artificial and natural grammars (Endress et al. 2009). For example, grammatical morphemes (e.g., the English plural “s”) occur in the first or the last position of words, but much more rarely in other positions (see general discussion for additional examples). However, little is known about the encoding of positional information in non-human animals. That is, there is substantial evidence that certain non-human primates and birds can encode the positions of items in a sequence (e.g., Chen et al. 1997; Hailman and Ficken 1987; Orlov et al. 2000, 2006; Terrace et al. 2003; Treichler et al. 2003). However, it is unknown whether positional information is encoded in a similar way as in humans, especially when tested under comparable conditions, and whether positional memories would dominate memory encoding also in non-human animals. Here, we start addressing these questions. Specifically, we contrast the capacity of chimpanzees (*Pan troglodytes*) and human adults to spontaneously encode the positions of items in sequences and to use this mechanism for extracting positional regularities from sequences. If they share this capacity, then it most likely evolved independently of language, and may not require language-input for developing ontogenetically in our own species.

We begin with a brief and selective review of the substantial literature on memory and sequence encoding and then turn to the evidence from our comparative experiments.

Two ways to encode sequences

Studies of memory encoding of sequences date back at least to Ebbinghaus (1885/1913). One important result of this research tradition is that sequences such as ABCD can be encoded in (at least) two different ways. On the one hand, it is possible to encode that A goes to B, B to C and C to D. Following Henson (1998), we will call such memories “chaining memories.” However, sequences such as ABCD

can also be encoded using a different mechanism, namely by remembering that A was in the first position, D in the last position and B in the second position. In other words, it is possible to link the items in the sequence to abstract positional codes (see, among many others, Conrad 1960; Henson 1998, 1999; Hicks et al. 1966; Ng and Maybery 2002; Schulz 1955). These codes are abstract because they are not bound to any specific sequence or sequence item. This is most apparent in so-called intrusion errors in memory experiments, where participants recall an element in its correct position—but in the wrong sequence (e.g., Conrad 1960). For example, following exposure to ABCD and EFGH, and when recalling the sequence EFGH, participants may erroneously recall the sequence EBGH; that is, the B item was erroneously included in the sequence, but it kept its correct position from its original sequence. As the B item was not “chained” to any of the items in the sequence EFGH, the positional codes must be sufficiently abstract to be generalized from one sequence to another. We call these kinds of memories “positional memories” and come back to their precise nature in the following text.

Artificial language learning experiments in humans and other animals have identified two similar sequence-learning mechanisms. Chaining memories are usually characterized in statistical terms as “transitional probabilities” (Aslin et al. 1998; Saffran et al. 1996). In a sequence ABCD, transitional probabilities reflect the conditional probabilities of A going to B, B to C and so on (rather than deterministic transitions as those studied in traditional memory research). While such computations were first demonstrated using continuous speech streams as stimuli, they work equally well on visual stimuli or musical tones (e.g., Fiser and Aslin 2002; Saffran et al. 1999). Also, the same mechanisms have been observed in a non-human primate (Hauser et al. 2001) and in rats (Toro and Trobalón 2005). They may thus reflect an evolutionarily ancient learning capacity.

Although less well studied, it has been shown that human adults can also acquire positional memories in the situation that is usually employed to investigate transitional probabilities, namely when participants are exposed to quasi-continuous speech (Endress and Bonatti 2007; Endress and Mehler 2009). In these experiments, participants learned that certain syllables had to occur word-initially or word-finally and generalized this regularity to new items they had never heard before. Much evidence, both from memory research and artificial grammar learning (e.g., Conrad 1960; Endress and Bonatti 2007; Endress and Mehler 2009; Henson 1998, 1999; Hicks et al. 1966; Ng and Maybery 2002; Schulz 1955), suggests that such memory for positions is distinct and independent from chaining memory. In Endress and Bonatti’s (2007) studies, for example, positional memories required different cues than chaining memories, seemed to dominate participants’

representations after little exposure (while chaining memories came to dominate after prolonged exposure), broke down under different conditions (Endress and Mehler 2009), and behaved differently under temporal reversal of the test items (that is, chaining memories worked equally well forward and backward, while positional memories broke down when the order of elements in the test items was reversed; Endress and Wood, in preparation; Turk-Browne and Scholl 2009). Moreover, phenomena such as the aforementioned intrusion errors are difficult to explain with chaining memories (e.g., Henson 1998, 1999). It thus seems reasonable to conclude that these two kinds of memories are indeed mediated by different mechanisms.

Previous experiments targeting serial learning abilities have revealed that different non-human species have some sensitivity to positional information (e.g., Chen et al. 1997; Hailman and Ficken 1987; Orlov et al. 2000, 2006; Terrace et al. 2003). In chick-a-dee calls, for example, certain note-types have to occur call-initially and others call-finally (Hailman and Ficken 1987), suggesting that these birds have a mechanism to track such positions. Likewise, Orlov et al. (2000) showed that macaque monkeys spontaneously link items to their sequential position. In each trial, the monkeys saw a sequence of visual shapes. Then, they saw all shapes of the sequence simultaneously together with a distracter shape on a touch screen and had to touch the shapes in the order in which they had previously seen them (without touching the distracter). Importantly, the distracter shapes were taken from other sequences the monkeys had seen. When the monkeys touched the distracter shapes, they tended to do so in sequential positions where the shape had occurred in its original sequence, suggesting that they had linked these items to their sequential positions. This pattern of errors is, therefore, reminiscent of the aforementioned intrusion errors in humans (e.g., Conrad 1960; Smith 1967).

How are sequential positions encoded?

Much memory research suggests that only edge positions may be encoded precisely, while all other positions appear to be encoded relative to the edges, and thus less precisely (e.g., Henson 1998). That is, according to most models of memory for sequential positions, items in a sequence become linked to edge-based markers, and their sequential position is derived from their distance to these marker points (e.g., Henson 1998; Hitch et al. 1996; Ng and Maybery 2002; Page and Norris 1998), even if the implementations in these models vary widely. It is important to note that the possibility that positions are encoded relative to sequence-edges cannot be reduced to a classic serial position effect (that is, the observation that items in sequence-edges are memorized better than items in

middles); rather, memory for positions seems to show its own serial position effect that is independent of the classic serial position effect in the Ebbinghaus tradition. This follows from the aforementioned observation that positional memory is distinct from other forms of sequential memory (e.g., Conrad 1960; Endress and Bonatti 2007; Endress and Mehler 2009; Henson 1998, 1999; Hicks et al. 1966; Ng and Maybery 2002; Schulz 1955). If so, serial position effects in these other forms of memory cannot explain the possibility that positional memory is edge-based. Rather, people seem to be endowed with specific, edge-based positional codes to which items in a sequence get linked; this allows them to reconstruct the sequential positions of items even if the items appear in a new sequence.

Further evidence for edge-based constraints comes from positional phenomena in artificial grammar learning experiments. While human adults can extract regularities involving the positions of items when the crucial positions are at the edges of sequences, they fail to do so when the crucial positions are in the middle of sequences (Endress et al. 2005; Endress and Mehler 2009). For example, learners notice that a syllable occurs in a particular position in a sequence when that position is at a sequence-edge, but have greater difficulty determining the position of a syllable when it is in another, non-edge position. Note that these results are not just due to the salience of the edges, as learners can process middle syllables perfectly well when they can rely on cues other than the positions (Endress et al. 2005).

In contrast to the aforementioned studies, results with long-tailed macaques seem to suggest that these animals encode sequential positions in absolute terms (that is, in terms of the first, second, third position and so on) rather than relative to the sequence-edges (Orlov et al. 2006). The basic paradigm in these experiments was similar to that used in Orlov et al.'s (2000), that is, the monkeys first saw a visual shape sequence and then had to touch the simultaneously presented shapes in the order in which they had been seen. In contrast to the previous experiments, however, monkeys were not shown the initial sample sequences but had to select the shapes according to their long-term memories of the previously trained sequences. Crucially, monkeys were trained on sequences of different lengths, namely three- and four-item sequences. This allowed the authors to contrast the predictions of absolute and relative encoding of the positions. For example, if positions were encoded relative to the sequence-edges, the proportion of intrusions of distracters in the last position of three-item sequences should be as high when the distracters are the last elements of other three-item sequences as when they are the last elements of other *four*-item sequences (because they would be the last elements in either case). In contrast, if positional encoding is absolute, the monkeys should be more likely to make intrusion errors if the distracter is the

last element of a three-item sequence than when it is the last element of a four-item sequence, because positions 3 and 4 are not equivalent in terms of absolute positions. Results showed that the monkeys were indeed less likely to touch distracters from four-item sequences when recalling three-item sequences than to touch distracters from sequences of the same length. Accordingly, Orlov et al. (2006) concluded that positional codes are absolute rather than relative.

There is an alternative interpretation of the Orlov experiments, one that directly connects with the experiments presented here: subjects may not only encode the sequential positions, but also the *length* of the sequences. If so, they may reject distracters from sequences of incorrect length, and one would expect the same pattern of results as that observed by Orlov et al. (2006). In fact, at least humans can remember in some circumstances (such as the tip-of-the-tongue experience) the length of words even when they cannot access the words (e.g., Brown and McNeill 1966; Koriat and Lieblich 1974). We also would argue that monkeys may plausibly encode the length of sequences when they have to learn them. Moreover, Orlov et al.'s (2006) data actually offer partial support for the relative encoding hypothesis. In their experiments, intrusions in incorrect positions (e.g., a distracter from position 2 that intrudes in position 3) are much more frequent in the second and third position than in the first and the last position (see their Figure 5A). This would be unexpected if positions were encoded absolutely. But the increased positional uncertainty in sequence-middles fits well with models encoding sequential positions relative to the sequence-edges. We thus believe that the current evidence is consistent with the idea that animals, humans included, encode sequential positions in relative terms. The following experiments attempt to provide additional support for this interpretation.

The current experiments

In the experiments reported in the following text, we presented chimpanzees (*Pan troglodytes*) and human adults with a situation in which they could encode both chaining regularities among items and regularities involving the positions of items, asking what kinds of information they extract from these sequences. While it is highly plausible that chimpanzees can process chaining dependencies among items, given that species as distant as humans, cotton-top tamarins and rats can do so (Hauser et al. 2001; Saffran et al. 1996; Toro and Trobalón 2005), it is less clear how they encode positions of items. Based on studies of human adults, one would expect that participants would initially encode information about an item's position, leaving for subsequent processing and additional exposure information about dependencies among items (Endress and

Bonatti 2007). This raises the following question: are humans particularly good at encoding positions in edges of sequences because many grammatical regularities are based on the positions of items, which, in turn, are encoded relative to the edges of different linguistic units? Humans may therefore know (consciously or unconsciously) from their extensive experience with language that edges constitute positions to "watch out" for. Another possibility (that is not necessarily incompatible with the first one) is that edge-based positional coding appears in other, non-linguistic domains, acting as a constraint on the structure of language. From this perspective, we might expect evidence for edge effects in non-human animals, a proposal that should not be overinterpreted. Specifically, we are not claiming that studies such as these will show that animals have language. In fact, our claim is almost exactly the opposite: that is, certain crucial properties of language might have non-linguistic origins that constrain the structure of language.

In the following experiments, we presented chimpanzees with materials that consisted of both positional and chaining regularities, and asked whether, given limited exposure, they were more likely to learn about edge-based positional information than about the dependencies among items. In other words, if chimpanzees follow the same general pattern as evidenced in human studies, then initially they should notice which items occur in the sequence-edges, while their sensitivity to other regularities should be much weaker.

Experiment 1: sequence learning in chimpanzees

Experiment 1 asks what kinds of information chimpanzees extract from sequences containing both positional and chaining regularities. Since these regularities are tracked by independent mechanisms in humans, chimpanzees may initially extract either of these regularities or both. More specifically, participants could learn that certain items occurred in the sequence-edges (i.e., a positional regularity) and that some items predicted others (i.e., a chaining regularity). To increase the chance of observing chaining information, we decided to make the items carrying the chaining information stand out in two ways. First, we used only three possible items in sequences of only six items, with the expectation that the limited number of items should make the chaining regularity easily detectable. Second, we set up the chaining regularity by associating the key items with more salient (i.e., in terms of acoustical dimensions such as pitch and amplitude) and functionally significant chimpanzee vocalizations than the item carrying the positional regularity.

Participants were first habituated to such sequences and then tested on new sequences that either respected or

violated the aforementioned chaining and positional regularities.

Materials and method

Participants

We tested 27 chimpanzees (20 females, mean age 5.37 years, range 1–21 years) from the Tchimpouanga sanctuary, Republic of Congo. This is a relatively naïve experimental population, but has been presented with various behavioral experiments over the last few years (Herrmann et al. 2007; Wood et al. 2007). Fifteen animals were included in the final analyses (seven adults, eight infants; see the following text for exclusion criteria). Approximately 1 year prior to this experiment, a subset of the present test subjects had been presented with some of the same tokens in a different habituation/discrimination paradigm (testing an AAB pattern; Hauser and Hare, in preparation). Thus, there was a long gap between the experiments, and minimal overlap in test subjects and test items. All subjects were born in the wild, lived in rich social and physical environments, and since an early age, they have been in close contact with human caretakers.

Stimuli

To explore the learning of chaining and positional regularities, we created sound sequences with three unique items: a pant grunt (X), a scream (A) and a copulation call (B), all recorded from wild chimpanzees unfamiliar to the test population. Both the chimpanzee scream and copulation call are very distinct from one another and are, we assume, both acoustically and functionally more salient than the grunt. Screams are emitted by a subordinate individual during moments of heightened aggression initiated by a dominant. Copulation calls occur, as indicated by their name, during copulation. Pant grunts, in contrast, are a more common occurrence, frequently emitted during relatively mild interactions between subordinates and dominants (e.g., Crockford and Boesch 2005). These items were combined into six item sequences with the stimuli arranged in different orders.

The average duration of the items was 240 ms ($X = 130$ ms, $A = 310$ ms, $B = 280$ ms, $SD = 96.4$). They were recorded to mono wav files with a sample rate of 44.1 kHz and a sample width of 16 bits. Sequences were created in Audacity (<http://audacity.sourceforge.net>) by pasting the items into new wav files (44.1 kHz, 16 bits, mono) in the order in which they were to appear in the sequences; items were separated by silences of 120 ms. Thus, the duration of each sequence was 1.71 s. Sequences were played at an intensity of 65–72 dB SPL.

Apparatus

The layout of the experiment is schematized in Fig. 1. Stimuli were played using an iPod Hi-fi speaker (Apple Inc., Cupertino, CA). Prior to testing, the speaker was placed out of the subject's sight and operated by the experimenter by means of a laptop computer. The experimenter was always positioned such that she was not in the line of sight between the subject and the speaker (see Fig. 1). The tests were also captured by a digital camcorder, which was operated by and positioned next to the experimenter.

Infant chimpanzees younger than 3 years were kept on a keeper's lap facing 180° away from the speaker. Animals of at least 3 years of age were tested alone in cages. Keepers directed subjects' attention away from the speaker by approximately 90°, either by playing with subjects or using food items; none of the keepers were aware of the goals or design of the experiments, and therefore were blind to our hypotheses. Furthermore, they were instructed not to react to the auditory stimuli.

Procedure

Subjects were tested individually using a habituation/discrimination paradigm. Sequences were played when an individual looked away from the speaker according to the coding criteria. Subjects were habituated to a series of auditory sequences that adhered to two patterns: (1) A preceded B and (2) X was located in the sequence-edges. There were four habituation sequences: XABXXX, XAXBXX, XAXXBX and XXXABX. These sequences were presented in random order until habituation was reached. We defined habituation as a failure to respond on three consecutive trials but also required subjects to hear a minimum of ten habituation trials. In cases where subjects produced three no-response trials before listening to ten habituation trials, we continued to present subjects with habituation trials until they had fulfilled both requirements.

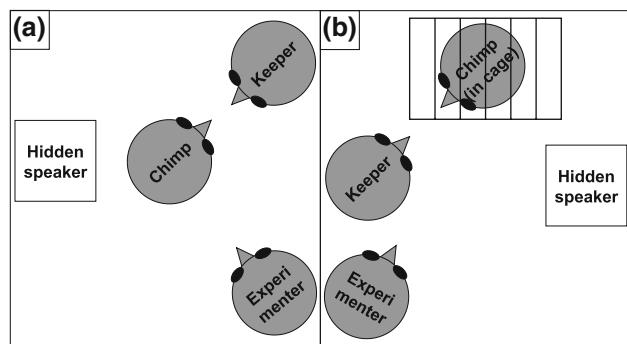


Fig. 1 Layout of Experiment 1 for infant subjects (a) and adult subjects (b)

Following habituation, we presented subjects with the six new test sequences indicated in Table 1. As we could not control the length of time subjects would remain attentive, we used a pseudorandom counterbalancing design. Specifically, the first four test trials consisted of two sequences that involved manipulations of the chaining regularity and two that involved manipulations of the positional regularity (sequences 1–4 in Table 1), respectively; these test sequences were presented in random order, and were then followed by two additional test sequences exploring the internal chaining relationships. As the data showed that participants kept responding at similar levels when reaching the last two sequences, these sequences were also included in the analyses.¹

Data acquisition and analysis

Upon completion of the experiment, SC blind-coded the final three habituation trials and all test trials, each in separate video clips with no sound; sound onset and offset were digitally flagged so that the response could be assessed relative to the playback. Due to the different positions of the speaker, different criteria were used to establish an orienting response in infants and adults, namely a turn toward the speaker of at least 90° or 45°, respectively. Trials not fulfilling these criteria were coded as non-responses. Trials were excluded from analysis if subjects left the camera view, were significantly distracted by another subject, or if there was a significant noise disturbance. We defined significant distraction as moments when the subject oriented toward or interacted with chimpanzees in neighboring areas. We defined noise disturbances as any sound (i.e., chimpanzee calls or the clanking of a cage) above the general background noise level. As the sounds were recorded using a camcorder, we could not use an objective criterion to define noise disturbances. (Camcorders automatically modify the recording gain depending on the general sound level; amplitudes in the recordings thus have no unique relation to the sound level in the environment.) While disturbances were thus defined subjectively, we observed only three such trials in the animals included in the final analysis, and two coders (SC and MH) agreed to exclude these three trials.

We assessed inter-observer reliability by having a second experienced coder (MH) analyze, blind to condition, a randomly chosen subset of 35 habituation and test trials. The two coders' judgments coincided on 34 of the 35 trials.

¹ If subjects had decreased their rate of responses in the last two trials, they should have responded less in these two trials than on earlier trials with the same properties, namely to items respecting the edge regularity and manipulating the chaining regularity. This, however, was not the case, $P = 0.48$ (Fisher's exact).

All χ^2 values are corrected for continuity. N values given with χ^2 values reflect the cumulative number of trials.

Subjects were excluded from analysis if, after blind-coding, it turned out that they did not have three non-responses in a row prior to starting the test phase or if they did not respond to any sequence at all during the test phase.

Results

On average, participants required 16.6 trials to habituate ($SD = 6.0$, range 10–27). As shown in Fig. 2 and Table 1, evidence of successful discrimination from the habituation material is revealed by a relatively stronger response to sequences violating the positional regularity (sequences 1 and 2 in Table 1; percentage of trials responded to: 46.7%) than to sequences respecting it (sequences 3–6 in Table 1; 20.0%; $\chi^2_{1,N=80} = 5.14$, $P = 0.023$, $\varphi = 0.282$). In contrast, subjects did not respond more to sequences violating the chaining relation (sequences 2, 4 and 6 in Table 1; 33.3%) than to sequences respecting it (sequences 1, 3, 5 in Table 1; 26.3%; $\chi^2_{1,N=80} = 0.15$, $P = 0.696$, $\varphi = 0.071$, ns). Power analysis revealed that the failure to observe a sensitivity to the chaining regularity was not due to insufficient statistical power, as one would need at least 260 chimpanzees (corresponding to 1,560 responses) to achieve a power of 80%.

To analyze more fully our results, we submitted our data to a binomial ANOVA (Venables and Ripley 2002) with the factors regularity type (positional vs. chaining) and test item type (legal vs. foil). We obtained a main effect of test item type ($\chi^2_{1,N=160} = 4.86$, $P = 0.028$), but no main effect of regularity type ($\chi^2_{1,N=160} = 0$, $P > 0.999$, ns), nor an interaction between these factors ($\chi^2_{1,N=160} = 1.78$, $P = 0.181$, ns).

The non-significant interaction raises the question of whether we had sufficient statistical power to detect it. Unfortunately, we are unaware of a well-accepted method of power analysis for binomial ANOVAs. We thus evaluated the power of the interaction as a function of the number of chimpanzees in the following way. For each number of chimpanzees, we generated 10,000 sample experiments. In each sample experiment, we randomly generated the number of orientations to each test item as an independent sample drawn from a binomial distribution, using the response rates observed in our experiment as orientation probabilities. Then we submitted the data from the sample experiment to the same binomial ANOVA as our empirical data and counted the proportion of sample experiments for which the interaction was significant at the 0.05 level. The results of our simulations showed that the interaction was significant in at least 80% of the simulations with at least 64 chimpanzees.

To assess whether the keepers' behavior might have influenced that of the chimpanzees, SC also blind-coded the behavior of the keepers during the test trials, using the same

Table 1 Summary of the test item types used

| Number | Item | Regularity chaining | Respected positional | Proportion reacted to/rejected | | | |
|--------|--------|---------------------|----------------------|--------------------------------|---------|---------|--------|
| | | | | Exp. 1 | Exp. 2a | Exp. 2b | Exp. 3 |
| 1 | AXXXXB | Yes | No | 8/15 | 23/30 | 20/30 | 20/30 |
| 2 | BXXXXA | No | No | 6/15 | 25/30 | 21/30 | 23/30 |
| 3 | XXABXX | Yes | Yes | 2/15 | 4/30 | 4/30 | 9/30 |
| 4 | XXBAXX | No | Yes | 5/13 | 17/30 | 4/30 | 8/30 |
| 5 | XXAXBX | Yes | Yes | 1/11 | 8/30 | 4/30 | 7/30 |
| 6 | XXBXAX | No | Yes | 2/11 | 17/30 | 10/30 | 15/30 |

Participants were tested on six unheard sequences. These could violate both the chaining regularity (sequences 2, 4, and 6) and the positional regularity (sequences 1 and 2). The three rightmost columns give the numbers of participants reacting to a sequence (for the chimpanzee experiment) or rejecting a sequence (for the human experiments)

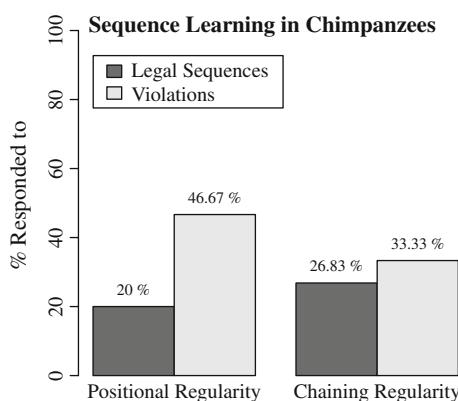


Fig. 2 Results of Experiment 1. Bars represent (from left to right) the percentage of dishabituations to sequences conforming to the positional regularity, violating it, conforming to the chaining regularity and violating it. While chimpanzees dishabituated more to sequences violating the positional regularity than to sequences conforming to it, there was no such difference for the chaining regularity

criteria as used with the chimpanzees. Out of a total of 64 trials (excluding those trials ($N = 26$) in which the keeper was not visible in the recording), the keepers oriented toward the speaker only on four occasions (6.25%). Their orientation rates differed neither depending on whether the test sequence respected the positional regularity ($\chi^2_{1,N=64} = 0.13$, $P = 0.724$, $\varphi = 0.026$, ns) nor depending on whether the test sequence conformed to the chaining regularity ($\chi^2_{1,N=80} = 0.34$, $P = 0.561$, $\varphi = 0.137$, ns). Crucially, in trials in which both the chimpanzees' and the keepers' behavior could be coded, the chimpanzees' orientation responses did not depend on whether the keepers oriented toward the speaker or not ($\chi^2_{1,N=58} = 0.10$, $P = 0.756$, $\varphi = 0.132$, ns).

For the adult chimpanzees, we also analyzed the trials separately where the subject was actually looking at the keeper. This was the case for 58.3% of the trials with adult subjects. However, as the keepers never oriented toward the speaker in these trials, it was not possible to evaluate

the association between the chimpanzees' behavior and that of the keepers. Together, these results indicate that the chimpanzees' responses to playbacks were independent of the keepers' behavior.

Discussion

The results of Experiment 1 suggest that when given the opportunity to learn either a chaining regularity, positional regularity or both, chimpanzees initially and spontaneously extracted the positional regularity under the test conditions presented. While the limited sample size used in Experiment 1 and, as a result, the limited statistical power do not allow for strong conclusions about the computational abilities of chimpanzees, it is striking that we found a reliable orienting response to violations of the positional regularity, but not to violations of the chaining regularity. We would argue that these results suggest that positional regularities are extracted more readily than chaining regularities, although we would ideally need more subjects to answer this question conclusively.

Before accepting this tentative conclusion, however, we need to rule out several alternative accounts of our data, relating to how the behavior of the keepers might have influenced that of the chimpanzees, whether the chimpanzees had equivalent exposure to both kinds of regularities and whether they simply might have attended only to the first and the last elements in the sequences.

Possibly, the chimpanzees might simply have reacted to cues provided by the keepers, without paying attention to the sequence at all, a possibility that is akin to the Clever Hans effect. In fact, although chimpanzees are not particularly attentive to human-provided cues except in competitive settings, they can follow human gaze (Bräuer et al. 2005, 2006). In Experiment 1, we minimized the possibility that subjects were just following human cues in several ways. First, the keepers were not informed about the experiment's goals and design and were specifically asked

not to react to the stimuli. Second, to rule out the possibility that they provided subconscious cues to the subjects, SC blind-coded their reactions during the test trials, using the same criteria as those used with the chimpanzees. Results showed that the keepers' behavior was not correlated with that of the chimpanzees. Moreover, in trials in which the (adult) chimpanzees actually looked at the keepers, the keepers never oriented toward the speaker, suggesting that they did not cue the chimpanzees either. Together, these results suggest that the chimpanzees' behavior was not based on cues provided by the keepers.

A second alternative interpretation of our results is that chimpanzees may simply have had more experience with the positional regularity than with the chaining regularity. In fact, all habituation items respected the positional regularity. The chaining regularity, in contrast, was implemented in a more variable way: in half of the habituation sequences, the A and B items were adjacent, while the remaining sequences had an intervening X item between the A and the B items. While this manipulation was necessary to avoid associating A and B items with any particular position within the sequences and to prevent participants from simply learning AB "chunks", it might have made it more difficult to learn the chaining regularity. For the moment, we leave open this possibility. In Experiment 3, however, we present data suggesting that it is unlikely that this design prevented chimpanzees from learning the chaining regularity.

A final alternative interpretation of these results is that the chimpanzees attended exclusively to the first (or the last) element of the sequence and ignored the rest. If so, they would dishabituate to sequences violating the chaining dependency not because they extracted the positional regularity, but because the only sequence element they attended to was "illegal" (that is, the first or the last item in the sequence). This possibility is unlikely for three reasons.

First, in a pilot study, the same chimpanzees attended to sequence-internal stimuli in a similar habituation/dishabituation paradigm (Hauser and Hare, in preparation). In particular, following habituation to chimpanzee vocalizations arranged in an AAB pattern, chimpanzee subjects were more likely to respond to novel sequences arranged in an ABB pattern than to novel sequences arranged in the familiar AAB pattern. To detect this difference, subjects must have noted either the change in the position of the identity relationship (i.e., from the first and second slot [AA] to the second and third slots [BB]) or the difference between identity (AA) and non-identity (AB). Either way, the chimpanzees were attending to more than the first or last element in these short, three-item strings.

Second, the sequence-internal vocalizations used during habituation were much more salient than those occurring in the sequence-edges, both acoustically and functionally; in

fact, it is highly unlikely that chimpanzees would ignore screams or copulation calls (that were used sequence-internally) rather than a pant grunt (that was used in the sequence-edges).

Third, given that all X items were physically identical, the A and B items were expected to "pop out" even though they were internal to the sequence. That is, these two call types were expected to stand out against the uniform background of X items. (The reader can verify this effect with an example sequence using human-produced sounds at the following address: <http://tinyurl.com/humanpopout>.) In vision, pop-out effects have been observed in humans (e.g., Treisman and Gelade 1980; Treisman and Gormican 1988), rhesus macaques (e.g., Bichot and Schall 1999) and chimpanzees (e.g., Tomonaga 1995). In audition, similar pop-out effects have been observed in humans (Cusack and Carlyon 2003). While we are not aware of any comparative work on auditory pop-out effects, they are highly likely to be shared by a wide variety of animals, given that they follow from general principles of auditory scene analysis and that these principles have been observed in many animals, including humans (Bregman 1990), Japanese macaques (Izumi 2002), European starlings (MacDougall-Shackleton et al. 1998), finches (Benney and Braaten 2000), bats (Moss and Surlykke 2001), and goldfish (Fay 1998, 2000). As a result, the A and B items are very likely to pop out, which would make it hard for the chimpanzees to ignore them.

We therefore suggest that it is unlikely that chimpanzees restricted their attention to the first (or the last) element of the sequences. More likely, the chimpanzees noticed that pant grunts occurred in the first, last or both positions, and when this positional regularity was violated, they responded by orienting to the speaker. Alternatively, they may have noticed that the scream and the copulation call did *not* occur in the sequence-edges during habituation and may have reacted to test sequences where these items were placed in the edges. In either case, after limited exposure, chimpanzees seem to have extracted a positional regularity rather than a chaining regularity.

Experiment 2: sequence learning in humans

Experiment 2 asked whether human adults would show the same pattern of responses as chimpanzees when tested under similar conditions. Thus, although certain aspects of the experimental design were necessarily different, we controlled for the amount of exposure and the particular details of the input to determine whether adult participants would preferentially detect violations at the edges, while being much less sensitive to violations concerning chaining regularities.

To make the materials used with human participants as comparable as possible to those used with chimpanzees, we

ran two different experimental conditions. In Experiment 2a, we used human-produced non-speech sounds of different saliency, such that the A and the B items were (presumably) more salient than the X item. However, while these items were human-produced, they were not human speech; in Experiment 2b, we test for the potential significance of this distinction by presenting human participants with human speech syllables.

Materials and method

Participants

Sixty native speakers of English (38 women, mean age 22.6 years, range 18–41) took part in this experiment. Half participated in Experiment 2a and half in Experiment 2b.

Stimuli

In Experiment 2a, X, A and B were a yawn, a belch and a scream, respectively, recorded from three different male humans. As the sounds differed in their subjective loudness after RMS normalization (due to different spectral content), we used Adobe Audition (Version 3.0) to manually adjust the amplitudes until the three sounds had roughly equal subjective loudness. X, A and B had a duration of 1,073, 567 and 516 ms, respectively.

In Experiment 2b, X, A and B were the syllables [faU], [hOI] and [SEI], respectively (in SAMPA notation). The syllables were synthesized using the us3 voice of mbrola (Dutoit et al. 1996) with a pitch of 150 Hz and a syllable duration of 400 ms.

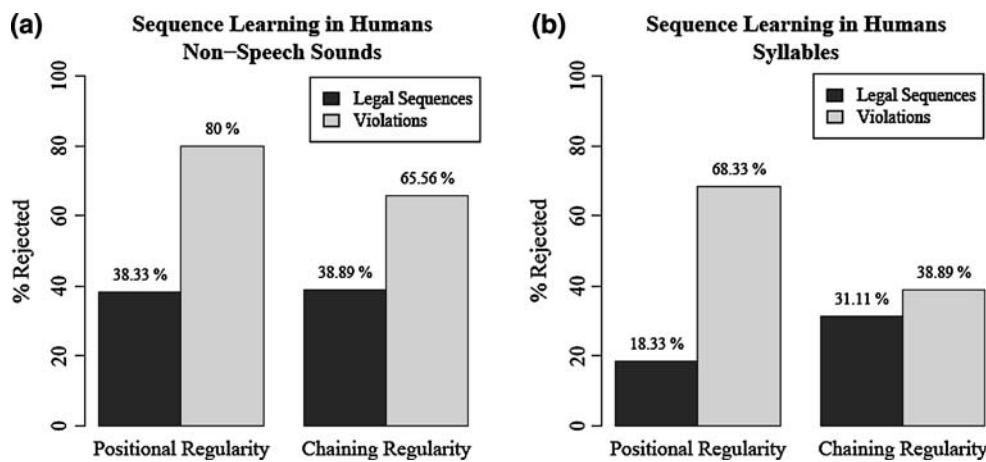


Fig. 3 Results of Experiment 2. **a, b** Bars represent (from left to right) the percentage of rejection of sequences conforming to the positional regularity, violating it, conforming to the chaining regularity and violating it. **a** Results of Experiment 2a. Participants were tested on human non-speech sounds. Human participants were more likely to reject sequences violating the positional regularity than those conforming it and more likely to reject sequences violating the

Apparatus

The experiment was run using PsyScope X software (<http://psyck.sissa.it>). Stimuli were presented over headphones; responses were collected from pre-marked keys on a keyboard.

Procedure

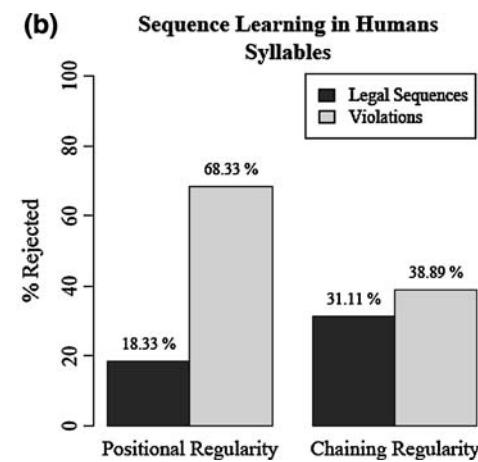
Participants were told that they would hear some sound sequences (in Experiment 2a) or a sequence of Martian words (in Experiment 2b). They were instructed to listen to these sequences/words. Then they were presented with the structurally identical familiarization sequences as the chimpanzees, each played four times; the number of presentations was thus the same as the mean number of trials to habituation in chimpanzees. Sequences were presented in random order; there was an inter-sequence interval of 1 s.

Following familiarization, we informed participants that they would hear six new sequences/words and that they would have to decide whether these were like the ones they just heard. We then presented them with the six test sequences (again using human-produced sounds instead of chimpanzee vocalizations); responses were indicated by pressing a key, revealing whether they thought the sequences were like the familiarization sequences or not.

Results

Experiment 2a

Figure 3a shows the results of Experiment 2a. Participants endorsed more sequences as being like the previous ones



chaining regularity than those conforming to it. The difference in rejection rates was higher for the positional regularity than for the chaining regularity. **b** Results of Experiment 2b. Participants were tested on human speech syllables. While human participants were more likely to reject the sequences violating the positional regularity than those conforming to it, there was no such difference for the chaining regularity

when these respected the positional regularity (endorsement percentage 61.67%) than when these violated the positional regularity (20.0%; $\chi^2_{1,N=180} = 26.19, P < 0.00001, \varphi = 0.393$). Their endorsement rates were also higher when the sequences respected the chaining regularity (61.11%) than when they did not (34.44%; $\chi^2_{1,N=180} = 11.78, P < 0.001, \varphi = 0.267$). A binomial ANOVA with the factors regularity type (positional vs. chaining) and test item type (legal vs. foil) yielded no main effect of regularity type ($\chi^2_{1,N=360} = 0, P > 0.999$), but a main effect of test item type ($\chi^2_{1,N=360} = 79.60, P < 0.00001$), and, crucially, an interaction between these factors ($\chi^2_{1,N=360} = 5.10, P = 0.024$).

Experiment 2b

The results of Experiment 2b are shown in Fig. 3b. Participants endorsed more sequences as being like the previous ones when these respected the positional regularity (81.7%) than when these violated the positional regularity (31.7%; $\chi^2_{1,N=180} = 41.79, P < 0.00001, \varphi = 0.494$). In contrast, participants did not show a difference in endorsement rates depending on whether the sequences respected the chaining regularity (68.9%) or not (61.1%; $\chi^2_{1,N=180} = 0.88, P = 0.344, \varphi = 0.081, \text{ns}$). A binomial ANOVA with the factors regularity type (positional vs. chaining) and test item type (legal vs. foil) yielded no main effect of regularity type ($\chi^2_{1,N=360} = 0, P > 0.999, \varphi = 0, \text{ns}$), but a main effect of test item type ($\chi^2_{1,N=360} = 28.61, P < 0.00001$) and an interaction between these factors ($\chi^2_{1,N=360} = 16.42, P < 0.00001$).

Discussion

The results of Experiments 2a and 2b suggest that, as the chimpanzees in Experiment 1, human adults predominantly extract positional information from sequences after short exposure. In Experiment 2a, where human-produced non-speech sounds were used, participants were significantly more sensitive to the positional regularity than to the chaining regularity, while they failed to generalize the chaining regularity in Experiment 2b, where speech syllables were used.

Participants in Experiment 2a (where human non-speech sounds were used) successfully discriminated sequences respecting the chaining dependency from sequences violating it. As in the chimpanzee experiment discussed earlier, learning the chaining regularity was likely facilitated by pop-out effects of the A and B items, respectively, because a belch (i.e., the A item) and a scream (i.e., the B item) are most likely to stand out against the background of yawns (i.e., the X item). Importantly, however, in both Experiment 2a and 2b, participants learned the positional regularity better than

the chaining regularity, suggesting that they preferentially encoded the positional regularity.²

These results suggest that positional regularities are easier to learn than chaining regularities. However, as mentioned earlier, chimpanzees and human adults may have performed better on the positional regularity because they had more experience with it than with the chaining regularity; indeed, the chaining regularity was implemented in a more variable way than the positional regularity, with half of the habituation sequences composed of adjacent A and B items, while the remaining sequences had an intervening X item between the A and the B items. If the participants' difficulties with the chaining regularity were due to a lack of experience with that regularity, they should succeed to generalize this regularity after more extensive exposure.

Experiment 3 was designed to test this possibility. Human adults in Experiment 2b, like chimpanzees in Experiment 1, failed to show a sensitivity to the chaining regularity. In Experiment 3, we replicated the general design of Experiment 2b, but increased the exposure fourfold. If the participants' difficulties with the chaining regularity were due to a lack of exposure to this regularity, they should succeed in Experiment 3.

Experiment 3: sequence learning in humans with more exposure

Materials and methods

Experiment 3 was identical to Experiment 2b (where human speech syllables were used), except that the

² There is another crucial difference between Experiment 2a and Experiments 1 (and Experiment 2b for that matter). Specifically, in Experiment 2a, the X item was almost twice as long as the A and B items. In Experiment 1, in contrast, the A and B items were almost twice as long as the X item (while there was no duration difference in Experiment 2b). This raises the possibility that human participants in Experiment 2a might have learned the chaining dependency not because they could rely on pop-out effects, but rather because of these differences in item duration. To control for this possibility, we replicated Experiment 2a, but using the chimpanzee vocalizations from Experiment 1. Results showed that participants endorsed more test sequences as being like the familiarization ones when they respected the positional regularity (64.2%) than when they violated the positional regularity (25.0%; $\chi^2_{1,N=180} = 23.0, P < 0.00001, \varphi = 0.369$). Their endorsement rates were also higher when the sequences respected the chaining regularity (60.0%) than when they did not (42.2%; $\chi^2_{1,N=180} = 5.0, P = 0.025, \varphi = 0.178$). A binomial ANOVA with the factors regularity type (positional vs. chaining) and test item type (legal vs. foil) yielded no main effect of regularity type ($\chi^2_{1,N=360} = 0, P > 0.999$), but a main effect of test item type ($\chi^2_{1,N=360} = 26.7, P < 0.000001$), and, crucially, an interaction between these factors ($\chi^2_{1,N=360} = 4.4, P = 0.037$). These results suggest that the differential results of Experiments 1 and 2a are not due to differences regarding the relative duration of the items in the sequences, but that learning the chaining regularity in Experiment 2a was helped by pop-out effects.

familiarization sequences were played 16 times (as opposed to four times in Experiment 2b). We tested 30 new native speakers of English (21 women, mean age 22.1, range 18–33) in this experiment.

Results and discussion

As shown in Fig. 4, participants endorsed more sequences as being like the previous ones when these respected the positional regularity (67.50%) than when these violated the positional regularity (28.33%; $\chi^2_{1,N=180} = 23.19$, $P < 0.00001$, $\varphi = 0.371$). In contrast, participants did not show a difference in endorsement rates depending on whether the sequences respected the chaining regularity (60.0%) or not (48.89%; $\chi^2_{1,N=180} = 1.81$, $P = 0.178$, $\varphi = 0.112$, ns). A binomial ANOVA with the factors regularity type (positional vs. chaining) and test item type (legal vs. foil) yielded no main effect of regularity type ($\chi^2_{1,N=360} = 0$, $P > 0.999$), but a main effect of test item type ($\chi^2_{1,N=360} = 20.41$, $P < 0.0001$) and an interaction between these factors ($\chi^2_{1,N=360} = 7.08$, $P = 0.0078$).

A binomial ANOVA with factors familiarization length (Experiment 2b vs. Experiment 3), regularity type (positional vs. chaining) and test item type (legal vs. foil) yielded a main effect of familiarization length ($\chi^2_{1,N=720} = 8.96$, $P = 0.003$), suggesting that participants were more likely to reject sequences in Experiment 3 than in Experiment 2b. We also obtained a main effect of test item type ($\chi^2_{1,N=720} = 47.90$, $P < 0.0001$), and, crucially, an interaction between test item type and regularity ($\chi^2_{1,N=720} = 22.42$, $P < 0.0001$). There were no other main effects or interactions. In other words, the main difference

between Experiments 2a and 3 was that participants were more likely to reject sequences after longer exposures. Crucially, however, their sensitivity to the chaining regularity did not improve despite a quadrupled familiarization duration.

General discussion

Memory and artificial language learning experiments have suggested that sequences can be memorized by two distinct kinds of mechanisms (e.g., Endress and Bonatti 2007; Henson 1998). One tracks *chaining* relations among items in a sequence, for example that one syllable predicts another one with a certain probability (e.g., Aslin et al. 1998; Saffran et al. 1996). The other mechanism tracks the *positions* of items in a sequence. That is, it memorizes which items occur in the first and the last position of sequences.

In this report, we compared the spontaneous performance of chimpanzees and human adults on a sequence-learning task where both positional and chaining regularities can be learned. In line with previous work with human adults, chimpanzees tracked the positional regularity after limited exposure to the sequences, but showed no sensitivity to the chaining regularity. Human adults showed a similar pattern of results when tested on human speech syllables. In contrast, when listening to human-produced non-speech sounds, human adults learned both the positional regularity and the chaining regularity, but the sensitivity to the positional regularity was much stronger. Under some circumstances, human adults, and possibly also chimpanzees, thus track chaining information. However, for both humans and chimpanzees, the initial encoding of sequences arises spontaneously and is dominated by positional information. Given that chimpanzees and humans were tested with similar methods and materials, and generated similar patterns of results, it is thus possible that the underlying mechanism is similar and highly sensitive to encoding positions in sequences.

Interestingly, edge-based positional regularities are frequent in language. Take morphology as an example. In English, one can add morphemes to the final edge of a word (as in appear-ed, where the /ed/ morpheme is added to a word stem to signal the past-tense) or to the leading edge of a word (such as dis-appear); morphemes, with a few exceptions, are not added in other positions. This is not specific to English: across the languages of the world, prefixes and suffixes are frequently used for grammatical purposes, while infixes (where morphemes are added to other positions than the edges) are rare (Greenberg 1957). The same is true for, say, stress assignment. Stressed syllables are either word-initial (as in Hungarian) or

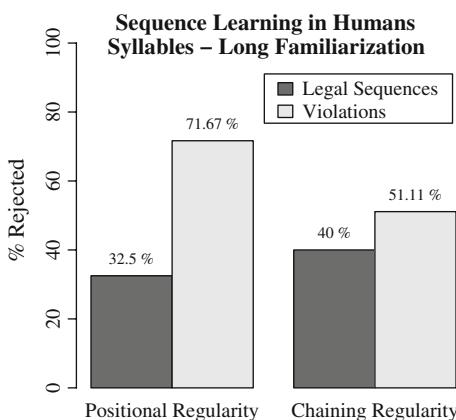


Fig. 4 Results of Experiment 3. Bars represent (from left to right) the percentage of rejection of sequences conforming to the positional regularity, violating it, conforming to the chaining regularity and violating it. While human participants were more likely to reject the sequences violating the positional regularity than those violating it, there was no such difference for the chaining regularity, even when given a fourfold familiarization relative to Experiment 2b

word-final (as in French) or at another position counted from one of the edges; no language assigns stress relative to other positions than the edges (e.g., Halle and Vergnaud 1987; Hayes 1995; Kager 1995). More generally, when grammatical regularities appeal to positions of items in sequences such as words, phrases or sentences, they tend to use the edges of these sequences as anchor points. For example, many linguistic regularities require that the edges of constituents on different levels of different linguistic hierarchies have to be aligned (e.g., McCarthy and Prince 1993; Nespor and Vogel 1986). In English, for instance, the onset of a sentence is also the onset of a word, whose onset, in turn, coincides with the onset of a morpheme. While this example may seem somewhat trivial, numerous complex linguistic regularities are typically formalized by assuming that edges of constituents on different levels have to be aligned (e.g., McCarthy and Prince 1993; Nespor and Vogel 1986). It is thus possible that these regularities appeal to an edge-based positional memory mechanism that is shared with chimpanzees, and thus, not specific to linguistic knowledge or competence.

This conclusion parallels previous discussions about the specificity of language mechanisms to humans. Although non-human animals clearly do not speak, some mechanisms used for speech perception may have predated its inception. For example, categorical perception of speech sounds, the perception of prototypical vowels and the compensation for co-articulated phonemes were all initially thought to be special to (human) speech (e.g., Eimas et al. 1971; Kuhl 1991; Liberman and Mattingly 1989, 1985; Mann 1986), but turned out to be shared with an array of other species (e.g., Kuhl and Miller 1975; Kluender et al. 1987; Kluender and Greenberg 1989; Lotto et al. 1997). Hence, while these capacities are recruited for clearly different purposes in humans and other animals, some basic underlying mechanisms must be shared. If the aforementioned linguistic regularities are due to an edge-based memory mechanism, a similar conclusion may hold for more grammatical aspects of language. While other animals almost certainly do not share our full-blown syntactic machinery (e.g., see Fitch and Hauser 2004; Gentner et al. 2006), at least some basic computational mechanisms such as the edge-based memory mechanism may be shared. That is, our results do not reveal linguistic capacities in non-human animals, but rather how non-linguistic capacities might be used for linguistic purposes. Such evolutionarily ancient mechanisms may thus explain why edge-based, positional regularities are learned particularly easily, and why such regularities appear to be a universal feature of human languages.

Acknowledgments The chimpanzee vocalizations were provided by K. Slocombe and S. Townsend. Funding for this experiment was

provided by MBB grants to M. Hauser, A. Endress, E. Versace and S. Carden, as well as additional funds to M. Hauser from the Wenner-Gren Foundation, McDonnell Foundation and gifts from J. Epstein and S. Shuman. We are grateful to L. Pharoah, R. Atencia, K. Brown and the Jane Goodall Institute USA and staff of Tchimpounga Sanctuary for their help and enthusiasm with our research. In particular, we appreciate the hard work of the animal caregivers: J. Maboto, B. Moumbaka, A. Sitou, M. Makaya, B. Bissafi, C. Ngoma, W. Bouity, J. Tchikaya, L. Bibimbou, A. Makosso, C. Boukindi, G. Nzaba, B. Ngoma. We also appreciate permission from the Congolese Ministère de la Recherche Scientifique et de l'Innovation Technique for allowing us to conduct our research in their country. We thank J. Call for helpful comments on an earlier draft of this manuscript. Lastly, we thank Brian Hare for his efforts in coordinating the logistical aspects of the work and for helping to make Tchimpounga a superb research facility.

References

- Aslin RN, Saffran J, Newport EL (1998) Computation of conditional probability statistics by 8-month-old infants. *Psychol Sci* 9:321–324
- Benney KS, Braaten RF (2000) Auditory scene analysis in estrildid finches (*Taeniopygia guttata* and *Lonchura striata domestica*): a species advantage for detection of conspecific song. *J Comp Psychol* 114:174–182
- Bichot NP, Schall JD (1999) Saccade target selection in macaque during feature and conjunction visual search. *Vis Neurosci* 16:81–89
- Bräuer J, Call J, Tomasello M (2005) All great ape species follow gaze to distant locations and around barriers. *J Comp Psychol* 119:145–154
- Bräuer J, Kaminski J, Riedel J, Call J, Tomasello M (2006) Making inferences about the location of hidden food: social dog, causal ape. *J Comp Psychol* 120:38–47
- Bregman AS (1990) Auditory scene analysis: the perceptual organization of sound. MIT Press, Cambridge
- Brown R, McNeill D (1966) The “tip of the tongue” phenomenon. *J Verb Learn Verb Behav* 5:325–337
- Byrne R (1999) Imitation without intentionality. Using string parsing to copy the organization of behaviour. *Anim Cogn* 2:63–72
- Chen S, Swartz K, Terrace HS (1997) Knowledge of the ordinal position of list items in rhesus monkeys. *Psychol Sci* 8:80–86
- Conrad R (1960) Serial order intrusions in immediate memory. *Br J Psychol* 51:45–48
- Conway CM, Christiansen MH (2001) Sequential learning in non-human primates. *Trends Cogn Sci* 5:539–546
- Crockford C, Boesch C (2005) Call combinations in wild chimpanzees. *Behaviour* 142:397–421
- Cusack R, Carlyon RP (2003) Perceptual asymmetries in audition. *J Exp Psychol Hum Percept Perform* 29:713–725
- Dutoit T, Pagel V, Pierret N, Bataille F, van der Vreken O (1996) The MBROLA project: towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In: Proceedings of the fourth international conference on spoken language processing, vol 3. Philadelphia, pp 1393–1396
- Ebbinghaus H (1885/1913) Memory: a contribution to experimental psychology. Teachers College, Columbia University, New York. <http://psychclassics.yorku.ca/Ebbinghaus/>
- Eimas P, Siqueland E, Jusczyk PW, Vigorito J (1971) Speech perception in infants. *Science* 171:303–306
- Endress AD, Bonatti LL (2007) Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition* 105:247–299

- Endress AD, Mehler J (2009) Primitive computations in speech processing. *Q J Exp Psychol* 62:2187–2209
- Endress AD, Scholl BJ, Mehler J (2005) The role of salience in the extraction of algebraic rules. *J Exp Psychol Gen* 34:406–419
- Endress AD, Nespor M, Mehler J (2009) Perceptual and memory constraints on language acquisition. *Trends Cogn Sci* 13:348–353
- Fay RR (1998) Auditory stream segregation in goldfish (*Carassius auratus*). *Hear Res* 120:69–76
- Fay RR (2000) Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*). *J Assoc Res Otolaryngol* 1:120–128
- Fiser J, Aslin RN (2002) Statistical learning of new visual feature combinations by infants. *Proc Natl Acad Sci USA* 99:15822–15826
- Fitch WT, Hauser MD (2004) Computational constraints on syntactic processing in a nonhuman primate. *Science* 303:377–380
- Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC (2006) Recursive syntactic pattern learning by songbirds. *Nature* 440:1204–1207
- Greenberg J (1957) Essays in linguistics. University of Chicago Press, Chicago
- Hailman J, Ficken M (1987) Combinatorial animal communication with computable syntax: chick-a-dee calling qualifies as ‘language’ by structural linguistics. *Anim Behav* 34:1899–1901
- Halle M, Vergnaud JR (1987) An essay on stress. MIT Press, Cambridge
- Hauser MD, Newport EL, Aslin RN (2001) Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* 78:B53–B64
- Hayes B (1995) Metrical stress theory: principles and case studies. University of Chicago Press, Chicago
- Henson R (1998) Short-term memory for serial order: the start-end model. *Cogn Psychol* 36:73–137
- Henson R (1999) Positional information in short-term memory: relative or absolute? *Mem Cogn* 27:915–927
- Herrmann E, Call J, Hernández-Lloreda MV, Hare B, Tomasello M (2007) Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science* 317:1360–1366
- Hicks R, Hakes D, Young R (1966) Generalization of serial position in rote serial learning. *J Exp Psychol* 71:916–917
- Hitch GJ, Burgess N, Towse JN, Culpin V (1996) Temporal grouping effects in immediate recall: a working memory analysis. *Q J Exp Psychol* 49:116–139
- Izumi A (2002) Auditory stream segregation in Japanese monkeys. *Cognition* 82:B113–B122
- Kager R (1995) Consequences of catalexis. In: van der Hulst H, van de Weijer J (eds) Leiden in last: Hil phonology papers i. Holland Academic Graphics, The Hague, pp 269–298
- Kluender KR, Greenberg S (1989) A specialization for speech perception? *Science* 244:1530
- Kluender KR, Diehl R, Killeen P (1987) Japanese quail can learn phonetic categories. *Science* 237:1195–1197
- Koriat A, Lieblich I (1974) What does a person in a ‘TOT’ state know that a person in a ‘don’t know’ state doesn’t know? *Mem Cogn* 2:647–655
- Kuhl PK (1991) Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Percept Psychophys* 50:93–107
- Kuhl PK, Miller JD (1975) Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* 190:69–72
- Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21:1–36
- Liberman AM, Mattingly IG (1989) A specialization for speech perception. *Science* 243:489–494
- Lotto A, Kluender KR, Holt L (1997) Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *J Acoust Soc Am* 102:1134–1140
- MacDougall-Shackleton SA, Hulse SH, Gentner TQ, White W (1998) Auditory scene analysis by European starlings (*Sturnus vulgaris*): perceptual segregation of tone sequences. *J Acoust Soc Am* 103:3581–3587
- Mann VA (1986) Distinguishing universal and language-dependent levels of speech perception: evidence from Japanese listeners’ perception of English “I” and “r”. *Cognition* 24:169–196
- McCarthy JJ, Prince A (1993) Generalized alignment. In: Booij G, van Marle J (eds) Yearbook of morphology 1993. Kluwer, Boston, pp 79–153
- Moss CF, Surlykke A (2001) Auditory scene analysis by echolocation in bats. *J Acoust Soc Am* 110:2207–2226
- Nespor M, Vogel I (1986) Prosodic phonology. Foris, Dordrecht
- Ng HL, Maybery MT (2002) Grouping in short-term verbal memory: is position coded temporally? *Q J Exp Psychol A* 55:391–424
- Orlov T, Yakovlev V, Hochstein S, Zohary E (2000) Macaque monkeys categorize images by their ordinal number. *Nature* 404:77–80
- Orlov T, Amit DJ, Yakovlev V, Zohary E, Hochstein S (2006) Memory of ordinal number categories in macaque monkeys. *J Cogn Neurosci* 18:399–417
- Page MP, Norris D (1998) The primacy model: a new model of immediate serial recall. *Psychol Rev* 105:761–781
- Prather JF, Peters S, Nowicki S, Mooney R (2008) Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* 451:305–310
- Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–1928
- Saffran JR, Johnson E, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. *Cognition* 70:27–52
- Schulz RW (1955) Generalization of serial position in rote serial learning. *J Exp Psychol* 49:267–272
- Smith K (1967) Rule-governed intrusions in the free recall of structured letter pairs. *J Exp Psychol* 73:162–164
- Suzuki R, Buck JR, Tyack PL (2006) Information entropy of humpback whale songs. *J Acoust Soc Am* 119:1849–1866
- Terrace HS (2005) The simultaneous chain: a new approach to serial learning. *Trends Cogn Sci* 9:202–210
- Terrace HS, Son LK, Brannon EM (2003) Serial expertise of rhesus macaques. *Psychol Sci* 14:66–73
- Tomonaga M (1995) Visual search by chimpanzees (Pan): assessment of controlling relations. *J Exp Anal Behav* 63:175–186
- Toro JM, Trobalón JB (2005) Statistical computations over a speech stream in a rodent. *Percept Psychophys* 67:867–875
- Treichler FR, Raghnati MA, Tilburg DNV (2003) Linking of serially ordered lists by macaque monkeys (*Macaca mulatta*): list position influences. *J Exp Psychol Anim Behav Process* 29:211–221
- Treisman A, Gelade G (1980) A feature-integration theory of attention. *Cogn Psychol* 12:97–136
- Treisman A, Gormican S (1988) Feature analysis in early vision: evidence from search asymmetries. *Psychol Res* 95:15–48
- Turk-Browne NB, Scholl BJ (2009) Flexible visual statistical learning: transfer across space and time. *J Exp Psychol Hum Percept Perform* 35:195–202
- Venables W, Ripley B (2002) Modern applied statistics with S, 4th edn. Springer, New York
- Wood JN, Glynn DD, Phillips BC, Hauser MD (2007) The perception of rational, goal-directed action in nonhuman primates. *Science* 317:1402–1405