

*Review*

# Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning

Carel ten Cate<sup>1,\*</sup> and Kazuo Okanoya<sup>2,\*</sup>

<sup>1</sup>*Behavioural Biology, Institute of Biology and Leiden Institute for Brain and Cognition, Leiden University, PO Box 9505, 2300 RA Leiden, The Netherlands*

<sup>2</sup>*Cognitive and Behavioural Science, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Meguro, Tokyo 153-8902, Japan*

The domain of syntax is seen as the core of the language faculty and as the most critical difference between animal vocalizations and language. We review evidence from spontaneously produced vocalizations as well as from perceptual experiments using artificial grammars to analyse animal syntactic abilities, i.e. abilities to produce and perceive patterns following abstract rules. Animal vocalizations consist of vocal units (elements) that are combined in a species-specific way to create higher order strings that in turn can be produced in different patterns. While these patterns differ between species, they have in common that they are no more complex than a probabilistic finite-state grammar. Experiments on the perception of artificial grammars confirm that animals can generalize and categorize vocal strings based on phonetic features. They also demonstrate that animals can learn about the co-occurrence of elements or learn simple ‘rules’ like attending to reduplications of units. However, these experiments do not provide strong evidence for an ability to detect abstract rules or rules beyond finite-state grammars. Nevertheless, considering the rather limited number of experiments and the difficulty to design experiments that unequivocally demonstrate more complex rule learning, the question of what animals are able to do remains open.

**Keywords:** birdsong; artificial grammar learning; syntax; language; vocalization; rule learning

## 1. INTRODUCTION

Human language is far more complexly structured than any other animal communication system. Using a limited set of vocal items—speech sounds—we are able to create numerous words, which can be combined into an infinite number of meaningful sentences using grammar. Although this level of complexity is beyond any animal vocal communication system known to date, many animal vocalizations show some form of structural organization, which is best characterized as a ‘phonological syntax’ [1]. Study of the abilities of animals to produce or perceive acoustic structures of a certain level of complexity is of particular relevance as it is especially the domain of syntax that is seen both as a core property of the language faculty and as the most critical difference between language and animal vocalizations [2]. Studying these abilities therefore not only reveals what animals are able to do, but can also provide insight on what might have been the precursors of language and on how linguistic complexity may have arisen. In this paper, we first concentrate on the structure of animal vocalizations. We discuss what animals use as ‘vocal units’ and how species-specific

vocalizations are structured, with emphasis on what is known about the regularities and rules underlying the production of song in songbirds. In the second part of the paper, we review experiments on the perceptual abilities of animals: are they able to detect, learn and generalize various ‘artificial grammar’ structures?

## 2. THE STRUCTURE OF ANIMAL VOCALIZATIONS

### (a) *Vocal variations and meaning*

The complexity of animal vocalizations varies from simple monosyllabic calls to complexly structured songs of birds (e.g. nightingales [3]) or whales (e.g. humpback whales [4]). Many animals have different vocalizations in their repertoires, which convey different messages: alarm calls for various predators (e.g. vervet monkeys [5]), calls related to food quality (e.g. ravens [6]) and vocalizations such as songs, related to mate choice or territorial defence. Variations within vocalizations can be due to motivational fluctuations [7] and can affect their meaning. For instance, altering the relative duration of the ‘trill’ and ‘flourish’ part of a chaffinch song affects the responses of male and female receivers [8]. This difference in ‘meaning’ is, however, of a different order compared with linguistic variation.

In spite of the large differences between language and animal vocalizations in use and meaning, they can nevertheless be compared when focusing on their

\* Authors for correspondence (c.j.ten.cate@biology.leidenuniv.nl; cokanoya@mail.ecc.u-tokyo.ac.jp).

One contribution of 13 to a Theme Issue ‘Pattern perception and computational complexity’.

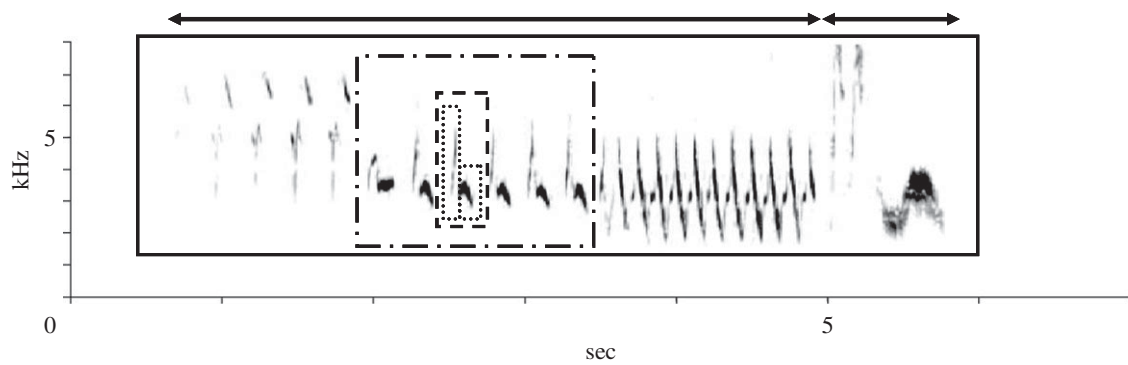


Figure 1. Sonogram of a chaffinch song, showing the different levels of segmentation. The smallest units (dotted lines) are single *elements*, separated by a brief gap. Two of these form one unit (the *syllable*, dashed line) that is repeated several times. A series of identical syllables form a *phrase*. In this song (enclosed by the unbroken line), there are three phrases consisting of brief, several times repeated syllables. This song ends, like many chaffinch songs, with a few elements that show a different spectrographic structure and are often not repeated. The first three phrases together form what is known as the ‘trill part’ of the chaffinch song (indicated by the left arrow), which is followed by what is known as the ‘flourish’ part (right arrow). (Sonogram courtesy of Katharina Riebel.)

structure: are they constructed from smaller units, and if so, what do these look like, how are they processed and what are the ‘syntactic rules’ that underlie their production? Here, we discuss what is known about the natural units of animal vocalizations, and how their phonological syntax can be analysed and extracted.

### (b) *Unit of production*

When we produce speech, this is made up of smaller units, phonemes. Each culture has a specific set of phonemes that include vowels and consonants. Each human language use less than 200 phonemes and the number tends to decrease, typically to less than 50 [9], for modern languages. Behavioural and linguistic studies indicate that our unit of speech production is a syllable, a combination of a vowel and optional consonants [10].

Do animals have a similar unit of vocal production? Many species can produce strings of several, differently structured, vocal units (‘elements’ or ‘syllables’), such as in the songs of many birds or whales [4,8]. The first major problem one encounters when describing the structure of such songs is to define the elementary building blocks. The basis for classifying units is usually how they appear on sonograms. An uninterrupted trace on a sonogram is typically taken as the smallest vocal unit (figure 1). However, separate traces on a sonogram need not coincide with units of production, as subsequent traces might always be produced together as a single unit. Therefore, additional information is needed to detect the elementary units. Indicators may be incidental, natural breaks in vocalizations, a variable number of repetitions of a unit, or varying pause durations between units [11]. Cynx [12] used an experimental approach to detect the units underlying the production of bird songs. He presented a singing zebra finch with a strobe light and determined when the bird stopped singing. Induced song termination mostly occurred after a song ‘syllable’, in this case defined as a continuous trace on the sonogram, and only rarely in the middle of a syllable. In another study, Franz & Goller [13] examined the pattern of expiration and inspiration during singing. This showed that visually

identified units usually corresponded to separate expiratory (and sometimes inspiratory) pressure pulses. Such methods help to get a better picture of the units of production, although they do not always lead to corresponding results [11]. Many more animal species also have vocalizations consisting of smaller units. Although these have not been as extensively studied as those of songbirds, similar experiments to those of Cynx have assessed that the separate traces on sonograms of the coos of collared doves [14], and the calls of tamarins [15] constitute separate units for production.

### (c) *Unit of learning*

Another window on the natural units of vocalizations is provided by comparing the vocalizations of tutors and pupils in vocal learning species. Zebra finches are songbirds, and, like other songbird species, learn their songs at an early age from a ‘tutor’, normally a conspecific male. Songbirds are not the only animals showing vocal learning. It is also present in parrots, hummingbirds, dolphins and whales, pinnipeds and bats [16]. However, most of our knowledge is based on studies of songbirds, and both the song learning process as well as the vocal complexity of their songs make songbirds currently the best model for comparative studies of language [17]. Song learning in songbirds occurs in two stages: one is sensory learning and the other is sensory motor learning. During sensory learning, auditory patterns of conspecific songs are memorized in the early juvenile period. During motor learning, vocal utterances are gradually shaped into songs by reference to the memorized auditory template [18].

What does the copying process indicate about the natural unit of songs? By examining songs of zebra finches exposed to two tutors in succession, ten Cate & Slater [19] showed that the finches chunked and segmented several syllables from tutor songs and recombined these in various orders. This was also found in another zebra finch study examining learning from several tutors [20]. It confirms that song fragments and not the song as a whole is the relevant entity. Although there is a clear tendency to copy chunks consisting of several units, when the songs of several pupils from the same

tutor are compared, one sees not only similarities in where they cut off the chunks, but also differences, with some birds leaving out or adding elements, in many cases coinciding with separate traces on a sonogram. These observations demonstrate that also in song learning smaller segments are the unit, but that these might be grouped into larger ones.

#### (d) *Animal phonetic categories?*

Having identified the smallest units in animal vocalizations, can these be allocated to categories, comparable to different human speech sounds? For non-learning species, this is usually not complicated. The between-individual variation is usually highly similar to the within-individual variation and the set of units is limited [14]. For vocal learners, this is more complex. Different individuals of a species may use elements that look different. Nevertheless, initially by inspection by eye [21,22], but nowadays by more advanced and objective clustering methods [23], it is possible to identify element categories. However, species, and also populations within species, can differ in the number and demarcation of these categories [23]. Such categories may represent the animal analogue of phonemes [24].

#### (e) *Extraction of phonological syntax from vocalizations*

As many animal vocalizations, learned or not, are built up from smaller units, this implies that there must be 'programmes' to arrange and produce the units in specific sequences. Animal species usually have recognizable species or population-specific patterns in element sequencing. For instance, the songs of individual chaffinches may differ in the type of elements they use, but they almost always are characterized by two to five series of repeated elements (trills) of decreasing frequency, followed by some differently structured, usually non-repeated, longer elements, the flourish (figure 1).

Several methods have been employed to extract the species-specific 'syntactic' structures from animal vocalizations, in particular from songbirds [11], but also for non-songbirds, such as hummingbirds [25] and whales [26]. Earlier attempts primarily used the transition diagram based on element-to-element transitional probabilities (known as a first-order Markov model or bigram) [27]. The weaknesses of such probabilistic representations involve their sensitivity to subtle fluctuations in datasets and their inability to address relationships more complex than adjacency. The  $n$ -gram model (also called an  $(n-1)$ th-order Markov model) attempts to overcome the second weakness by addressing probabilistic relationships longer than immediate adjacency and predicting the  $n$ -th elements based on the preceding  $n-1$  elements [28]. A variant of this model involves changing the length of  $n$  according to the data so that  $n$  is always optimal. The hidden Markov model (HMM) is among the most applicable models in terms of its explanatory power [11]. This model assumes that actual elements of a song represent output from hidden states [29,30]. The number of these hidden states, the transitional relationships between the

states and the probability with which each song element emerges from each of the states has to be estimated, which requires prior knowledge or assumptions about the nature of the states, as well as inferences about the transitional relationships between them [31]. For descriptive purpose, the  $n$ -th order Markov models and its variants are most appropriate [32]. On the other hand, the HMM can provide a powerful tool to infer neural mechanisms for song sequence generation [30,33]. However, all these models are finite-state grammars and the conclusion of studies of vocal structure in birds, including those of species with extensive and elaborate singing styles such as starlings [32,34] and nightingales [3], is that neither their vocal complexity nor that of any other animal species studied to date extends beyond that of a probabilistic finite-state grammar [35] (also see Hurford [36] for further discussion).

The presence of species specificity in structure of many species' songs does not imply that the syntax develops free of experience. While this can be so, as in many non-songbirds (e.g. collared doves [37]), several examples in songbirds show that species-universal patterns of vocal structure can also be due to learning. For instance, nightingales have a repertoire of up to 200 distinctive songs. Normally nightingales hardly ever repeat a song immediately—a 'universal' syntax pattern. This pattern is known as 'immediate variety', in contrast to the 'eventual variety' shown by a bird like the chaffinch [38], which usually repeats a particular song a number of times before switching to a new type. However, when Hultsch [39] exposed young nightingales to strings of songs showing an eventual variety pattern, these birds later on showed considerable song repetition, demonstrating a learning effect on this pattern. Another example concerns the sequencing of different units within a song in white-crowned sparrows. Songs of this species usually start with a whistle, followed by a buzz and a trill. In tutoring experiments, Soha & Marler [40] showed that when young white-crowned sparrows are exposed to separate elements, they have the tendency to tie these together, starting with a whistle element. While this suggests a predetermined universal song structure, a later experiment by Rose *et al.* [41] indicated a role for experience. When units of two elements were presented in the reverse order from the usual sequence, the birds combined them and produced a reversed element sequence, instead of the normal species-specific one. Song syntax in birds thus can show some flexibility and be affected by what they heard as juveniles.

#### (f) *Bengalese finch songs: an integrative example*

Song syntax is extensively studied in Bengalese finches [42]. Bengalese finches sing complex songs with variable song note transitions. Several notes together can form chunks and multiple chunks are sung in variable, but more or less determined orders. The formal complexity of Bengalese finch songs is at most finite-state syntax [35]. With this simple rule-based song structure, it is possible to study behavioural determinants for song chunking and segmentation.

First, the unit of song production was determined using the same procedure as applied by Cynx [12]. When a strobe light was flashed to the singing Bengalese finch, song termination occurred more often at the point where the transitional probability between notes was low but not where it was high [43]. Thus, song termination seldom occurred within a song chunk, but typically occurred at the edge of the chunk. Next, the developmental process for the song note chunking was examined in a semi-natural rearing experiment. Adult male and female Bengalese finches were kept in a large aviary, where they raised a total of 40 male chicks [44]. When the chicks reached adulthood, which part of the song came from which tutor was examined. Note-to-note transition probabilities were analysed using second-order transition analysis. Approximately 80 per cent of the Bengalese finches learned from between two and four tutors. The results imply the operation of three underlying processes. First, juvenile finches segmented the continuous singing of an adult bird into smaller units. Second, these units were produced as chunks when the juveniles practised singing. Third, juvenile birds recombined the chunks to create an original song. As a result, the chunks copied by juveniles had higher transition probabilities and shorter intervals of silence than those characterizing the boundaries of the original chunks. These processes suggest that Bengalese finches segmented songs by using both statistical and prosodic cues, such as pause duration, during learning.

Finally, an analysis of Bengalese finch song strings was done by using a set of procedures for automatically producing a deterministic finite-state automaton [28]. Based on an  $n$ -gram representation, song elements were chunked to yield a hierarchically higher order unit. The transitions among these units were mapped and processed for  $k$ -reversibility, where  $k$ -reversibility referred to the property of the resulting automaton that was able to determine the state that existed  $k$  steps back from the present state [45]. This set of procedures provided a robust estimation of automaton topography and has been useful in evaluating the effects of developmental or experimental manipulations of birdsong syntax. The analysis indicated no need to assume a higher level of formal syntax than a finite-state one to describe the phonological syntax in Bengalese finch songs [35]. Combining song analysis with neurobiological studies of the same system [46,47] provided also an understanding of the neural mechanisms of Bengalese finch vocal complexity.

### 3. THE STRUCTURE OF ANIMAL VOCALIZATIONS: DISCUSSION

Although animal species may differ tremendously in the structure and complexity of their vocalizations, they all seem to have a few features in common. In all species studied to date, the production of vocalizations occurs in smaller units—‘elements’ or ‘syllables’ that often can be identified as separate traces on a sonogram. These seem to be the minimal units of production, which are combined to generate more complex species-specific vocalizations. This species specificity does, however, not imply that the rules underlying production emerge

without experience. Detailed studies of songbirds, such as the Bengalese finch, are beginning to reveal the neural basis of the underlying grammar. However, in spite of the enormous variation among animal vocalizations, there are currently no examples of structures more complex than a finite-state grammar [35].

### 4. ANIMAL PERCEPTION OF SYNTAX

While there is no evidence of animals being able to produce vocal structures of greater complexity than a probabilistic finite-state grammar [35], this need not imply that they cannot detect greater complexity in vocal input. So, can animals detect particular grammatical patterns and learn the abstract rules involved? Over the last 10 years, addressing this question has been inspired by studies on human subjects using ‘artificial grammar learning’ (AGL) paradigms [48]. In AGL studies, human subjects are exposed to artificially created meaningless strings of sounds, e.g. nonsense words consisting of consonant–vowel syllable combinations, structured according to a particular algorithm. After exposure to a variety of strings, the subjects are tested with novel strings not heard before, which either fit the training algorithm or not. The responses to the test strings reveal what the subjects learned about the underlying algorithm. This method is not only used with human adults, but is also widely applied to examine the grammar detection abilities of prelinguistic human infants. It is also excellently suited for comparative studies addressing these abilities in animals.

The mechanisms involved in AGL in humans need not map one-to-one with the mechanisms underlying natural language learning or be specific to the language domain [48]. Nevertheless, experiments using this paradigm indicate the presence or absence of cognitive abilities to abstract rules from specific input. This feature is a characteristic for language and examining which level of abstraction animals are able to achieve gives a clue about similarities and differences between humans and animals. Below, we summarize findings obtained in animal studies using various AGL tasks. While these studies shed light on the rule detection abilities of animals, they also make clear that there are a lot of open questions and that the search for the syntactic abilities of animals has only just begun.

#### (a) Statistical learning

In a seminal paper, Saffran *et al.* [49] exposed eight-month-old infants to three-syllable ‘words’, e.g. ‘bidaku’. Three such words were presented as a continuous stream, generated with a speech synthesizer, e.g. *bidakupadotigolabidaku...* The only cues to word boundaries were the transitional probabilities between syllable pairs, which were high within words (1.0, e.g. *bida*) and low between words (0.33 in all cases, e.g. *kupa*). The infants were first presented with the speech stream for 2 min. Subsequently, in the test phase, they were presented with lists of either the statistical words or ‘non-words’ consisting of three syllables in a non-familiar order. In another experiment, statistical words were contrasted with part-words consisting of syllables that were paired with low probability during the familiarization. Infants



looked longer to both non-words and part-words than to the statistical words, hence detected a difference. Various studies confirmed and extended these findings, for both infants and human adults. They also showed that non-speech input, such as musical tones [50] can similarly be segmented based on transitional probabilities. Furthermore, adults can detect some structures based on non-adjacent dependencies, i.e. when the first and third speech segments (consisting of consonants or vowels) are a fixed combination, with the second one less predictable [51].

The above experiments inspired several studies aimed at testing whether the observed ability is specific to humans or can be observed in non-human animals as well. Hauser *et al.* [52] exposed cotton-top tamarins to the same strings as used by Saffran *et al.* [49]. After exposure to the speech stream, the response to the test sequences was measured by examining the orientation towards a speaker playing the test sounds. The tamarins showed more responding towards novel than towards familiar sequences, indicating that they detected the difference.

Toro & Trobalon [53] showed that rats could also segment the speech stream. This experiment used a different type of habituation design: rats were first trained to press a lever for food in the presence of the speech stream. After reaching a stable performance, it was examined how exposure to non-words, words and part-words affected the lever-pressing response. The study showed that rats' lever-pressing was affected by the stimulus type, but rather than being sensitive to transitional probabilities, the rats responded to the frequency of co-occurrence of syllables.

The Newport & Aslin [51] study on the infant abilities to detect non-adjacent dependencies by statistical learning has also been replicated using animals. Newport *et al.* [54] showed that tamarins were able to detect non-adjacent dependencies, although there were some stimulus-dependent differences between humans and tamarins. Toro & Trobalon [53] tested whether rats could detect such dependencies, but obtained no evidence for this.

### (b) *Learning a specific grammar*

Several studies [48,55–57] exposed young infants to artificially constructed grammars in which a limited set of items ('words'; e.g. PEL, VOT, TAM, BIFF, CAV) was arranged in strings ('sentences') according to specific finite-state grammars. These grammars could generate strings of varying lengths and with varying transitional probabilities between the items. After exposure to a limited number of strings generated by a specific grammar, infants were able to discriminate novel sequences generated with a 'predictive' grammar, in which positions of a particular item had a predictive value for the next one, from sequences generated with either a non-predictive grammar or with another predictive grammar, or random sequences [56]. They also could detect the grammar structure when each position could be occupied by more than one item [57]. These experiments suggest that infants developed an abstract notion of the grammar beyond specific word order. However, a caveat in experiments in which the same items are used for training and test strings is that infants

may have been using the transitional probabilities to discriminate various patterns from each other, rather than acquired knowledge of the underlying grammar structure [58]. A stronger test was provided by Gomez & Gerken [55] who showed that infants can distinguish new strings produced by their training grammar from strings produced by another grammar despite a change in vocabulary between training and test.

Saffran *et al.* [57] presented the same stimulus sets they used for infants also to tamarins. These could distinguish novel strings with the predictive grammar from non-grammatical strings, but not when more than one item occupied a specific position in the sequence. Recently, Abe & Watanabe [59] replicated the experiment by Saffran *et al.* [57] in Bengalese finches, replacing the speech syllables by song elements. The finches were exposed to one set of strings and a change in calling rate when hearing the novel strings was used to assess whether the birds differentiated between the strings. Similar to infants and tamarins, the Bengalese finches exposed to the predictive strings distinguished novel strings from this grammar from those generated with the non-predictive grammar, whereas Bengalese finches exposed to non-predictive grammar did not make the distinction.

Herbranson & Shimp used a different approach to examine grammar learning in pigeons [60,61]. The pigeons were trained in an operant design, in which they were presented with strings of coloured letters (T,X,V,P,S) on a monitor, and taught to peck one key for one type of sequence and another for the other type. Strings organized according to finite-state grammars comparable to those used by Gomez & Gerken [55] and Saffran *et al.* [57] were tested against strings with the same items either in a non-grammatical order or, in another experiment, according to a different grammar. After a substantial training period, the pigeons were able to make the distinction between the two sets in each experiment, albeit using a modest learning criterion (60% correct). When tested with novel strings of the training grammars, some birds transferred the discrimination to these. However, although it is clear that the pigeons learned some regularities present in the grammar, a detailed analysis of similarities between training and test strings revealed that they responded to specific substrings shared between training and testing stimuli. Hence, they classified test items, at least partially, according to which extent they shared transitional probabilities between the items with the training stimuli. Therefore, it is not clear whether and what the pigeons really learned about the underlying grammar [60].

The stimuli used in the pigeon experiment were visual ones and the birds could observe the full sequence rather than, as in experiments using vocal items, being exposed to the items one by one. This might reduce comparability with the experiments on tamarins and Bengalese finches, which used vocal strings. Nevertheless, the pigeon study points at a problem also recognized for vocal strings [57], namely that the often limited number of strings presented in training and tests makes it difficult to rule out the possibility that lower level statistical regularities present in these strings, such as transitional probabilities

between units, give rise to the observed results. Even though some potential cues might be ruled out, it is difficult to control for all possible regularities, with the risk of obtaining accidental 'false positives'.

### (c) *Learning 'algebraic' rules*

Marcus *et al.* [58] used a different approach to examine rule learning in infants. In a classic study, seven-month-old infants were exposed to three-syllable sequences following either an ABA (e.g. gatiga, linali, talata, etc.) or an ABB (e.g. gatiti, linana, talala) pattern. When subsequently presented with strings using novel syllables, either consistent or inconsistent with the familiar structure, they were able to distinguish these, apparently detecting the underlying string structure. To test whether it was reduplication of syllables ('titi', 'nana', etc.) that was used to discriminate the sentences, another group of infants was trained and tested with AAB versus ABB strings, again showing the ability to distinguish the sequences. Another test controlled for phonetic cues that might be used to distinguish between the strings. From the experiment, Marcus *et al.* [58] propose that infants 'extract algebra-like rules that represent relationships between placeholders such as 'the first item X is the same as the third item Y''. A subsequent experiment showed that even five-month-old infants can make the distinction, provided they are trained and tested with congruent combinations of vocal and visual stimuli [62].

This experiment too was repeated with tamarins. However, the study that claimed to have demonstrated that tamarins could distinguish an AAB from an ABB pattern [63] has recently been retracted after an internal examination at Harvard University 'found that the data do not support the reported findings' [64]. Clearly this experiment needs independent replication. Another study [65] showed that rhesus monkeys can distinguish AAB from ABB sequences. In this study, 'A's and 'B's were different call types. Individuals were habituated with up to 10 unique strings (i.e. consisting of different calls, but all arranged according to the same rule) and next tested with four novel strings: two in one pattern and two in the other. The rhesus monkeys showed a stronger dishabituation to test strings following the non-familiar pattern.

A different procedure to examine rule learning was used for rats [66]. In a classical conditioning design, rats were exposed to isolated stimuli of either an XYX (ABA and BAB), an XXY (AAB and BBA) or a YXX (ABB and BAA) structure and received a food pellet after presentation of one stimulus type (e.g. XYX), but not after the other two (e.g. XXY and YXX). The response time of the rats to the different stimuli was used as an indicator for detecting a difference. The stimuli in this case were light/dark sequences with each element lasting 10 s. The rats managed to discriminate the sets, indicating a sensitivity to the stimulus sequences. In a second experiment, rats received food after an XYX stimulus but not after an XXY or YXX stimulus, with the stimuli consisting of pure tones. Next, they were tested with novel tone strings, transposed to a different

frequency. The rats differentiated in their response time according to the pattern. The experiment thus showed generalization to novel stimuli, but it also raises several questions. Apart from the duration and nature of the stimuli being quite unlike a vocal string, and the question of whether the novel test sequence is sufficiently different to demonstrate a clear transfer [67], another issue concerns the question of what exactly the rats have learned. They might discriminate the XYX from XXY and YXX strings by using some algebraic rule in the sense as suggested by Marcus *et al.* [58], e.g. that first and last items of the stimulus should be identical (for the positive stimuli) or different (for the negative ones). But they might also attend to the presence or absence of element reduplications, or whether subsequent elements are always distinct, or may have learned the individual items by rote memorization. Although some of these alternative explanations are more plausible than others [66,67], lower level explanations should be excluded by further experiments before we can conclude that the rats learned an algebraic rule. For instance, if the rats attend to similarity of the first and last items of the sequence, they should treat a stimulus like XXYX as a negative one and XYYX as positive.

A recent experiment using zebra finches aimed at testing such alternatives [68]. The finches were trained to peck a key which triggered either an XYX stimulus (ABA or BAB—stimuli composed of zebra finch song elements, each bird had a different element type as A and B), an XXY or a YXX (AAB, ABB, BAA and BBA) stimulus. When hearing an XYX stimulus, they could obtain food by pecking a second key. If they pecked upon hearing XXY or YXX, the light in the cage was switched off for some seconds. The zebra finches mastered this go-nogo task, thus demonstrating sensitivity to the training sequences. In subsequent tests, one out of eight birds generalized the distinction to XYX sequences in which the B-element was replaced by a novel type C (i.e. the novel stimuli were CAC and ACA); the others, however, did not. Nevertheless, for all birds, the question remained how they distinguished the sequences. Using various (non-reinforced) test strings, the data suggest that the zebra finches attended to the presence of repeated elements and used this, rather than a rule of the kind suggested by Marcus *et al.* [58]. This finding is of interest as more recently Endress *et al.* [69,70] showed that humans also seem to be particularly sensitive to the presence of repetition-based structures, especially when they appear at the 'edges' (beginning or end) of strings. This may indicate that, in addition to statistical regularities and predispositions, language acquisition is guided by the so-called 'perceptual or memory primitives' [69,70], which might also be present or have precursors in animals.

### (d) *Learning centre-embedded hierarchical structures*

Studies addressing whether adult humans, infants or animals are able to distinguish hierarchical structures from non-recursive serial structures have attracted

considerable attention. Embedded structures are of higher complexity than the finite-state grammars used in the experiments discussed so far. One possible test for the ability to cope with full hierarchical structures, using formal language theory, was introduced by Fitch & Hauser [71]. In this study, adult humans and tamarins were both habituated to speech syllables arranged either in ABAB or ABABAB strings ((AB)<sup>n</sup>, a finite-state arrangement) or as AABB or AAABBB strings (A<sup>n</sup>B<sup>n</sup>, which requires a more powerful system beyond finite-state capabilities, termed 'phrase structure' grammars). Humans clearly distinguished both string types, irrespective of the string they were habituated with. However, only those tamarins habituated with the finite-state strings detected the difference from the other string type. When habituated with the phrase structure grammar, they failed to do so.

Although the experiment was presented as a test for abilities beyond the finite-state level, it gave rise to a lot of discussion about whether it was a proper test for recursion [72]. The study also inspired similar experiments on songbirds, respectively starlings [73] and zebra finches [74]. The stimuli used in these studies were conspecific song elements, which might be more salient to the subjects and hence provide a better chance of detecting whether birds can distinguish the different patterns. Both studies used a go-nogo design and trained the birds with  $n = 2$  strings, i.e. ABAB versus AABB. For the starlings, the 'A's and 'B's in the strings consisted of 'rattle' and 'warble' phrases taken from a starling song. For the zebra finches, these were different song elements. In both studies, each individual was trained with multiple strings, i.e. with different exemplars of 'A' and 'B' phrases or elements, or with the same 'A's and 'B's in different positions. After acquiring the distinction between the training sets, all birds were next exposed to novel strings, constructed from both novel song phrases and elements belonging to the same categories as the earlier ones, or familiar phrases and elements in a novel sequence. Both species showed a clear transition of the discrimination to the novel strings. However, this transfer to novel exemplars of a familiar type need not be based on acquired abstract knowledge of the structures, but might be brought about by attending to shared acoustic characteristics between the familiar and novel 'A's and 'B's [74]. Such phonetic generalization has been demonstrated for zebra finches distinguishing human speech sounds [75] and is a factor that has also been demonstrated to affect the outcome in AGL experiments in humans [76]. For this reason, the zebra finch study subsequently tested how the birds responded when A and B elements were replaced by elements of a different element type ('C's and 'D's). Only one bird showed a transfer of the discrimination to the novel sounds, suggesting that the earlier generalization to the novel 'A' and 'B' stimuli was based on shared phonetic cues. Even so, the question remained of what, if any, rules the birds used to discriminate the two string sets. This was examined for both starlings and zebra finches by presenting them with non-reinforced probes. In both cases, probes consisted of stimulus sequences with higher  $n$ -values (three or four),

irregular sequences with different numbers of 'A's and 'B's, and other arrangements of 'A's and 'B's (e.g. ABBA). Gentner *et al.* [73] concluded that their probe testing revealed that the starlings mastered the distinction of training sounds by being sensitive to the underlying A<sup>n</sup>B<sup>n</sup> grammar. This conclusion was challenged by van Heijningen *et al.* [74], who showed that the most likely 'rule' used by the zebra finches was attending to the presence or absence of repeated elements. They based this conclusion on examining the test results for each finch separately, which revealed individual differences in strategy, as also demonstrated in humans [77]. Based on this and other arguments, van Heijningen *et al.* [74] argue that the question of whether birds are able to detect a hierarchical context-free structure is still open (see Gentner *et al.* [78] and ten Cate *et al.* [79] for an exchange of views on the issue).

It should be noted that all the experiments described above lacked a critical feature of a centre-embedded structure [80], in which a specific 'A' is always linked to a specific 'B' (i.e. in a connection like '...if, ...then...'). Such linkage was present in a recent experiment by Abe & Watanabe [59]. In a habituation design, Bengalese finches were exposed to an artificial 'centre-embedded language' in which the presence of a specific element early in the strings predicted the presence and position of another specific element later in the strings. When tested with novel strings, either of the embedded language or deviations, they discriminated between the two. However, as test strings of the embedded language shared many more element sequences with the training strings than the deviating ones, it cannot be excluded that the Bengalese finches used acoustic similarity rather than abstract structure to distinguish the strings [81]. Thus, more research is necessary to test for this ability in animals.

## 5. SYNTAX DETECTION IN ANIMALS: DISCUSSION

### (a) *Can animals detect syntax patterns?*

Having reviewed different types of AGL experiments, what can we conclude on the abilities of animals to detect structure in vocal input?

A first thing to note is that animals are definitely capable of, and skilled in, detecting and learning about regularities in sequential vocal input. They can detect that two sequences constructed from the same items are different and detect differences in co-occurrence between items. They also are good at generalizing at the phonetic level, detecting phonetic regularities among, or differences between, items and strings and classifying novel ones on the basis of their similarities to known ones. What is less clear is what level of more abstract rule learning they are capable of. In several experiments, the tasks facing the animals might have been solved by using lower level strategies, such as attending to the phonetic similarities of items [73], to co-occurrence of items rather than transitional probabilities [53] or to presence or absence of reduplications of items (in relation to their position), rather than by detecting and using relational properties of items in different positions, algebraic rules [58] or mastering an



artificial grammar [57]. While this does not exclude the possibility that some animals learned abstract regularities, the use of simpler strategies remains a distinct possibility. In addition, while the aim of the animal experiments was to examine whether they are able to detect abstract relationships between items in the form of detecting ‘algebraic rules’, ‘non-adjacent dependencies’ or ‘phrase structure grammar’ as present in language, the results suggest no ability beyond simple finite-state grammar detection. Furthermore, a characteristic of language is that abstract operations are based on equivalence relating to functional categories. An ABA structure is, for instance, formed by a noun–verb–noun series, in which ‘A’ and ‘B’ stand for different categories, and in which the first A and the last one in the ABA structure usually are not the same, but share a functional similarity, e.g. both being nouns. In the animal experiments, researchers have, at best, used acoustic categories, in which the phonetic structure of one A shared its overall structure with that of another A. In the latter case, the abstraction, if present, is likely to be pattern-based and can, at best, be described in terms of relational operations linked to the physical structure of the tokens. This contrasts with the category-based generalization of the noun–verb–noun type, which is further removed from the physical identity of the tokens [55,69,82] and more characteristic of syntactic structure. So, taking all the above considerations together, the conclusion has to be that we are still in need of unambiguous demonstrations of animals able to detect structures based on higher order syntactical processing of strings, and that it remains a challenge to design meaningful and critical experiments.

The considerations raised above do not mean that the animal experiments are less valuable. They provide valuable insights in the cognitive abilities that animals have. Also, it must be emphasized that absence of proof of more abstract rule learning in this case certainly is no proof of absence: it is just very hard to design experiments that fully rule out simpler solutions. In addition, some of the comments apply equally to some AGL experiments in humans. Nevertheless, they have greatly advanced our insights by examining several basic computational skills that might be at the developmental or evolutionary basis of language. Therefore, designing animal experiments comparable to human ones still provides a powerful tool for comparative studies to get insight in the skills and mechanisms shared between humans and other animals.

### (b) Methodology

It should be noticed that various animal experiments used different methodologies. One aspect concerns the nature of the items used in the strings. In the tamarin experiments, these were human speech syllables, often identical to the ones used in experiments with human subjects. On the one hand, this increases the comparability of the experiments. On the other hand, however, one may argue that just like humans might be more attentive to speech sounds than other ones [83], or that such sounds might be processed differently, animals might be more attentive or sensitive

to vocal items constructed from conspecific sounds. As a consequence, they might be more likely to detect, or respond to, regularities in strings consisting of such items than of speech sounds.

A second methodological aspect concerns the procedure. Human AGL experiments usually use a short exposure period to one type of strings (positive examples only), followed by a short test phase applying a dishabituation paradigm (in infants), or asking subjects to classify novel items as belonging to a particular category. Some animal experiments, most notably those on tamarins, used a similar design. The other animal studies used a variety of different types of tasks based on various forms of classical [66] or operant [73] conditioning. There are various pros and cons of these methods. The habituation design has the advantage that it uses brief exposures, which it can be argued reflect the ‘spontaneous’ detection of structure in strings, compared with, at the other end, a go–nogo design that requires an extensive training phase. Also, in a habituation design, the animals are exposed to one type of strings and anything they learn is due to exposure to this single string type. In contrast, a go–nogo design always uses two categories of strings, differing in structure. This makes it harder to see whether there are asymmetries in learning about different string types, as are revealed in several habituation experiments. The go–nogo procedure may also encourage animals to attend to contrasts that may differ from the features of strings they would attend to spontaneously. On the other hand, if an animal shows no response when a novel stimulus type is presented after a habituation phase, it need not imply that the animal did not detect any difference—it may have done so, but did not show it in a behavioural change. This risk of a ‘false negative’ is less in a go–nogo design, where attending to stimulus differences has direct consequences for the animal in terms of reward or punishment. For the same reason, go–nogo experiments are very suited to examining the limits of the cognitive abilities of animals, because a task can be designed in such a way (at least in theory) that it ‘forces’ the animal to reveal whether it has the, normally hidden, ability to resolve it. Also, go–nogo or other conditioning designs may allow for the use of a more extensive set of test strings than is possible in a habituation design, again because the reward structure keeps the animal focused on responding to differences in sounds. To diminish the difference with the usual habituation tasks used in AGL experiments in humans, these too can be tested with a go–nogo design [75]. With such pros and cons, in combination with the difficulty to design unambiguous experiments in general, it may be the best strategy to test animals of a particular species using different paradigms. If these provide converging outcomes, it is all the more likely that the species really has a particular ability.

A methodological point of concern is the repeated use of the same grammar as that of Saffran *et al.* [57], in several experiments, in combination with presenting each individual with the same training and test strings. This introduces the problem of ‘pseudoreplication’ [84,85]. If any animal species shows the ability to



master one particular finite-state grammar, this would be a valid demonstration that animals can detect that grammar. However, to demonstrate that animals can detect the structure of finite-state grammars *in general*, other variants should also be used. In addition, several studies tested all individuals with identical strings. This bears the risk that these strings are not recognized by abstract structure, but by, for instance, a particular salient sound combination that by chance provides a phonetic cue. Therefore, if the goal is to demonstrate perception of an abstract pattern, this pattern should be instantiated in different ways for different individuals (e.g. different terminal units for the A and B categories).

## 6. CONCLUSION

We studied evidence from the spontaneously produced vocalizations of animals and from perceptual experiments using artificial grammars to examine to what extent or level animal syntactic computational abilities compare to those of humans using language.

With respect to production, many animals have, similar to humans, vocalizations consisting of a variety of vocal units (notes or elements) that are combined to create strings (chunks and phrases) that in turn can be produced in different patterns. However, while the patterns in animals differ between species and may also differ among individuals within species, they all have in common that the structures are no more complex than a probabilistic finite-state grammar [35].

As for perception of syntactic structures, the experiments so far mainly demonstrate and confirm the ability of animals to generalize and categorize vocal items based on phonetic features, or co-occurrence of items, but do not provide strong evidence for an ability to detect rules or patterns based on relationships beyond these or beyond simple 'rules' like attending to reduplications of units or the presence of particular units at the beginning or end of vocal stimuli. However, considering the still rather limited number of experiments and the difficulty of designing experiments that unequivocally demonstrate more complex rule learning, we are only scratching the surface, and hence the question of what animals are capable of remains open.

We thank Katharina Riebel, Claartje Levelt, Tecumseh Fitch and the referees for their comments on earlier versions of this paper.

## REFERENCES

- Marler, P. 2000 Origins of music and speech: Insights from animals. In *The origins of music* (eds N. Wallin, B. Merker & S. Brown), pp. 31–48. Cambridge, MA: MIT Press.
- Hauser, M. D., Chomsky, N. & Fitch, W. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)
- Todt, D. & Hultsch, H. 1998 How songbirds deal with large amount of serial information: retrieval rules suggest a hierarchical song memory. *Biol. Cybern.* **79**, 487–500. (doi:10.1007/s004220050498)
- Green, S. R., Mercado III, E., Pack, A. A. & Herman, L. M. 2011 Recurring patterns in the songs of humpback whales (*Megaptera novaeanghiae*). *Behav. Process.* **86**, 284–294. (doi:10.1016/j.beproc.2010.12.014)
- Seyfarth, R. M., Cheney, D. L. & Marler, P. 1980 Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim. Behav.* **28**, 1070–1094. (doi:10.1016/S0003-3472(80)80097-2)
- Bugnyar, T., Kijne, M. & Kotrschal, K. 2001 Food calling in ravens: are yells referential signals? *Anim. Behav.* **61**, 949–958. (doi:10.1006/anbe.2000.1668)
- Dabelsteen, T. & Pedersen, S. 1990 Song and information about aggressive responses of blackbirds, *Turdus merula*: evidence from interactive playback experiments with territory owners. *Anim. Behav.* **40**, 1158–1168. (doi:10.1016/S0003-3472(05)80182-4)
- Leitao, A. & Riebel, K. 2003 Are good ornaments bad armaments? Male chaffinch perception of songs with varying flourish length. *Anim. Behav.* **66**, 161–167. (doi:10.1006/anbe.2003.2167)
- Atkinson, Q. D. 2011 Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* **332**, 346–349. (doi:10.1126/science.1199295)
- Wijnen, F. 1988 Spontaneous word fragmentations in children: evidence for the syllable as a unit in speech production. *J. Phonet.* **16**, 187–202.
- ten Cate, C., Lachlan, R. F. & Zuidema, W. In press. Analyzing the structure of bird vocalizations and language: finding common ground. In *Birdsong, speech and language. Exploring the evolution of mind and Brain* (eds J. J. Bolhuis & M. Everaert). Cambridge, MA: MIT Press.
- Cynx, J. 1990 Experimental determination of a unit of song production in the zebra finch (*Taeniopygia guttata*). *J. Comp. Psychol.* **104**, 3–10. (doi:10.1037/0735-7036.104.1.3)
- Franz, M. & Goller, F. 2002 Respiratory units of motor production and song imitation in the zebra finch. *J. Neurobiol.* **51**, 129–141. (doi:10.1002/neu.10043)
- ten Cate, C. & Ballintijn, M. R. 1996 Dove coos and flashed lights: interruptibility of song in a nonsongbird. *J. Comp. Psychol.* **110**, 267–275. (doi:10.1037/0735-7036.110.3.267)
- Miller, C. T., Flusberg, S. & Hauser, M. D. 2003 Interruptibility of long call production in tamarins: implications for vocal control. *J. Exp. Biol.* **206**, 2629–2639. (doi:10.1242/jeb.00458)
- Janik, V. & Slater, P. 1997 Vocal learning in mammals. *Adv. Study Behav.* **26**, 59–99. (doi:10.1016/S0065-3454(08)60377-0)
- Bolhuis, J. J., Okanoya, K. & Scharff, C. 2010 Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* **11**, 747–759. (doi:10.1038/nrn2931)
- Doupe, A. & Kuhl, P. K. 1999 Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631. (doi:10.1146/annurev.neuro.22.1.567)
- ten Cate, C. & Slater, P. J. B. 1991 Song learning in zebra finches: how are elements from two tutors integrated? *Anim. Behav.* **42**, 150–152. (doi:10.1016/S0003-3472(05)80617-7)
- Williams, H. & Staples, K. 1992 Syllable chunking in zebra finch (*Taeniopygia guttata*) song. *J. Comp. Psychol.* **106**, 278–286. (doi:10.1037/0735-7036.106.3.278)
- Marler, P. & Pickert, R. 1984 Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza georgiana*). *Anim. Behav.* **32**, 673–689. (doi:10.1016/S0003-3472(84)80143-8)
- Zann, R. 1993 Variation in song structure within and among populations of Australian zebra finches. *Auk* **110**, 716–726.

- 23 Lachlan, R. F., Verhagen, L., Peters, S. & ten Cate, C. 2010 Are there species-universal categories in bird song phonology and syntax? A comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taenopygia guttata*), and swamp sparrows (*Melospiza georgiana*). *J. Comp. Psychol.* **124**, 92–108. (doi:10.1037/a0016996)
- 24 Yip, M. J. 2006 The search for phonology in other species. *Trends Cogn. Sci.* **10**, 442–446. (doi:10.1016/j.tics.2006.08.001)
- 25 Ficken, S. M., Rusch, K. M., Taylor, S. J. & Powers, D. R. 2000 Blue-throated Hummingbird song: a pinnacle of nonoscine vocalizations. *Auk* **117**, 120–128. (doi:10.1642/0004-8038(2000)117[0120:BTHSAP]2.0.CO;2)
- 26 Suzuki, R., Buck, J. R. & Tyack, P. L. 2006 Information entropy of humpback whale song. *J. Acoust. Soc. Am.* **119**, 1849–1866. (doi:10.1121/1.2161827)
- 27 Dobson, C. W. & Lemon, R. E. 1979 Markov sequences in songs of American thrushes. *Behaviour* **68**, 86–105. (doi:10.1163/156853979X00250)
- 28 Kakishita, Y., Sasahara, K., Nishino, T., Takahasi, M. & Okanoya, K. 2009 Ethological data mining: an automata-based approach to extract behavioral units and rules. *Data Min. Knowl. Disc.* **18**, 446–471. (doi:10.1007/s10618-008-0122-1)
- 29 Kogan, J. & Margoliash, D. 1998 Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study. *J. Acoust. Soc. Am.* **103**, 2185–2196. (doi:10.1121/1.421364)
- 30 Katahira, K., Suzuki, K., Okanoya, K. & Okada, M. 2011 Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. *PLoS ONE* **6**, e24516. (doi:10.1371/journal.pone.0024516)
- 31 Katahira, K., Okanoya, K. & Okada, M. 2007 A neural network model for generating complex birdsong syntax. *Biol. Cybern.* **97**, 441–448. (doi:10.1007/s00422-007-0184-y)
- 32 Gentner, T. Q. 2007 Mechanisms of temporal auditory pattern recognition in songbirds. *Lang. Learn. Dev.* **3**, 157–178. (doi:10.1080/15475440701225477)
- 33 Jin, D. Z. & Kozhevnikov, A. A. 2011 A compact statistical model of the song syntax in Bengalese finch. *PLoS Comp. Biol.* **7**, e1001108. (doi:10.1371/journal.pcbi.1001108)
- 34 Eens, M. 1997 Understanding the complex song of the European starling: an integrated ethological approach. *Adv. Study Behav.* **26**, 355–434. (doi:10.1016/S0065-3454(08)60384-8)
- 35 Berwick, R. C., Okanoya, K., Beckers, G. J. L. & Bolhuis, J. J. 2011 Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* **16**, 113–121. (doi:10.1016/j.tics.2011.01.002)
- 36 Hurford, J. R. 2012 *The origins of grammar*. Oxford, UK: Oxford University Press.
- 37 Ballintijn, M. R. & ten Cate, C. 1997 Vocal development and its differentiation in a non-songbird: the collared dove (*Streptopelia decaocto*). *Behaviour* **134**, 595–621. (doi:10.1163/156853997X00548)
- 38 Slater, P. J. B. 1983 Sequences of song in chaffinches. *Anim. Behav.* **31**, 272–281. (doi:10.1016/S0003-3472(83)80197-3)
- 39 Hultsch, H. 1991 Early experience can modify singing styles: evidence from experiments with nightingales, *Luscinia megarhynchos*. *Anim. Behav.* **42**, 883–889. (doi:10.1016/S0003-3472(05)80140-X)
- 40 Soha, J. A. & Marler, P. 2001 Vocal syntax development in the white-crowned sparrow (*Zonotrichia leucophrys*). *J. Comp. Psychol.* **115**, 172–180. (doi:10.1037/0735-7036.115.2.172)
- 41 Rose, G. J., Goller, F., Gritton, H. J., Plamondon, S. L., Baugh, A. T. & Cooper, B. G. 2004 Species-typical songs in white-crowned sparrows tutored with only phrase pairs. *Nature* **432**, 753–758. (doi:10.1038/nature02992)
- 42 Okanoya, K. 2004 Song syntax in Bengalese finches: proximate and ultimate analyses. *Adv. Study Behav.* **34**, 297–346. (doi:10.1016/S0065-3454(04)34008-8)
- 43 Seki, Y., Suzuki, K., Takahasi, M. & Okanoya, K. 2008 Song motor control organizes acoustic patterns on two levels in Bengalese finches (*Lonchura striata* var. *domestica*). *J. Comp. Physiol. A* **194**, 533–543. (doi:10.1007/s00359-008-0328-0)
- 44 Takahasi, M., Yamada, H. & Okanoya, K. 2010 Statistical and prosodic cues for song segmentation learning by Bengalese finches (*Lonchura striata* var. *domestica*). *Ethology* **116**, 481–489. (doi:10.1111/j.1439-0310.2010.01772.x)
- 45 Angluin, D. 1982 Inference of reversible languages. *J. ACM* **29**, 741–765. (doi:10.1145/322326.322334)
- 46 Nishikawa, J., Okada, M. & Okanoya, K. 2008 Population coding of song element sequence in the Bengalese finch HVC. *Eur. J. Neurosci.* **27**, 3273–3283. (doi:10.1111/j.1460-9568.2008.06291.x)
- 47 Fujimoto, H., Hasegawa, T. & Watanabe, D. 2011 Neural coding of syntactic structure in learned vocalizations in the songbird. *J. Neurosci.* **31**, 10 023–10 033. (doi:10.1523/JNEUROSCI.1606-11.2011)
- 48 Gomez, R. L. & Gerken, L. A. 2000 Infant artificial language learning and language acquisition. *Trends Cogn. Sci.* **4**, 178–186. (doi:10.1016/S1364-6613(00)01467-4)
- 49 Saffran, J. R., Aslin, R. N. & Newport, E. L. 1996 Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928. (doi:10.1126/science.274.5294.1926)
- 50 Saffran, J. R., Johnson, E. K., Aslin, R. N. & Newport, E. L. 1999 Statistical learning of tone sequences by human infants and adults. *Cognition* **70**, 27–52. (doi:10.1016/S0010-0277(98)00075-4)
- 51 Newport, E. L. & Aslin, R. N. 2004 Learning at a distance. I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* **48**, 127–162. (doi:10.1016/S0010-0285(03)00128-2)
- 52 Hauser, M. D., Newport, E. L. & Aslin, R. N. 2001 Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* **78**, B53–B64. (doi:10.1016/S0010-0277(00)00132-3)
- 53 Toro, J. M. & Trobalon, J. B. 2005 Statistical computations over a speech stream in a rodent. *Atten. Percept. Psychophys.* **67**, 867–875. (doi:10.3758/BF03193539)
- 54 Newport, E. L., Hauser, M. D., Spaepen, G. & Aslin, R. N. 2004 Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cogn. Psychol.* **49**, 85–117. (doi:10.1016/j.cogpsych.2003.12.002)
- 55 Gomez, R. L. & Gerken, L. A. 1999 Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* **70**, 109–135. (doi:10.1016/S0010-0277(99)00003-7)
- 56 Saffran, J. R. 2002 Constraints on statistical language learning. *J. Mem. Lang.* **47**, 172–196. (doi:10.1006/jmla.2001.2839)
- 57 Saffran, J. R., Hauser, M., Seibel, R., Kapfhammer, J., Tsao, F. & Cushman, F. 2008 Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* **107**, 479–500. (doi:10.1016/j.cognition.2007.10.010)
- 58 Marcus, G. F., Vijayan, S., Bandi Rao, S. & Vishton, P. M. 1999 Rule learning by seven-month-old infants. *Science* **283**, 77–80. (doi:10.1126/science.283.5398.77)
- 59 Abe, K. & Watanabe, D. 2011 Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* **14**, 1067–1074. (doi:10.1038/nn.2869)

- 60 Herbranson, W. T. & Shimp, C. P. 2008 Artificial grammar learning in pigeons. *Learn. Behav.* **36**, 116–137. (doi:10.3758/LB.36.2.116)
- 61 Herbranson, W. T. & Shimp, C. P. 2003 Artificial grammar learning in pigeons: a preliminary analysis. *Learn. Behav.* **31**, 98–106. (doi:10.3758/BF03195973)
- 62 Frank, M. C., Slemmer, J. A., Marcus, G. F. & Johnson, S. P. 2009 Information from multiple modalities helps five-month-olds learn abstract rules. *Dev. Sci.* **12**, 504–509. (doi:10.1111/j.1467-7687.2008.00794.x)
- 63 Hauser, M. D., Weiss, D. & Marcus, G. 2002 Rule learning by cotton-top tamarins. *Cognition* **86**, B15–B22. (doi:10.1016/S0010-0277(02)00139-7)
- 64 2010 Retraction notice. *Cognition* **117**, 106. (doi:10.1016/j.cognition.2010.08.013)
- 65 Hauser, M. D. & Glynn, D. 2009 Can free-ranging rhesus monkeys (*Macaca mulatta*) extract artificially created rules comprised of natural vocalizations? *J. Comp. Psychol.* **123**, 161–167. (doi:10.1037/a0015584)
- 66 Murphy, R. A., Mondragon, E. & Murphy, V. A. 2008 Rule learning by rats. *Science* **319**, 1849–1851. (doi:10.1126/science.1151564)
- 67 Corballis, M. C. 2009 Do rats learn rules? *Anim. Behav.* **78**, e1–e2. (doi:10.1016/j.anbehav.2009.05.001)
- 68 van Heijningen, C., Chen, J., van Laatum, I., van der Hulst, B. & ten Cate, C. Submitted. Rule learning by zebra finches in an artificial language learning task: which rule?
- 69 Endress, A. D., Scholl, B. J. & Mehler, J. 2005 The role of salience in the extraction of algebraic rules. *J. Exp. Psychol. Gen.* **134**, 406. (doi:10.1037/0096-3445.134.3.406)
- 70 Endress, A. D., Nespore, M. & Mehler, J. 2009 Perceptual and memory constraints on language acquisition. *Trends Cogn. Sci.* **13**, 348–353. (doi:10.1016/j.tics.2009.05.005)
- 71 Fitch, W. & Hauser, M. D. 2004 Computational constraints on syntactic processing in a nonhuman primate. *Science* **303**, 377–380. (doi:10.1126/science.1089401)
- 72 Fitch, W. T., Friederici, A. D. & Hagoort, P. 2012 Pattern perception and computational complexity: introduction to the special issue. *Phil Trans. R. Soc B* **367**, 1925–1932. (doi:10.1098/rstb.2012.0099)
- 73 Gentner, T. Q., Fenn, K. M., Margoliash, D. & Nusbaum, H. C. 2006 Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204–1207. (doi:10.1038/nature04675)
- 74 van Heijningen, C. A. A., de Visser, J., Zuidema, W. & ten Cate, C. 2009 Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl Acad. Sci. USA* **106**, 20 538–20 543. (doi:10.1073/pnas.0908113106)
- 75 Ohms, V. R., Gill, A., Van Heijningen, C. A. A., Beckers, G. J. L. & ten Cate, C. 2010 Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proc. R. Soc. B* **277**, 1003–1009. (doi:10.1098/rspb.2009.1788)
- 76 Onnis, L., Monaghan, P., Richmond, K. & Chater, N. 2005 Phonology impacts segmentation in online speech processing. *J. Mem. Lang.* **53**, 225–237. (doi:10.1016/j.jml.2005.02.011)
- 77 Zimmerer, V. C., Cowell, P. E. & Varley, R. A. 2010 Individual behavior in learning of an artificial grammar. *Mem. Cogn.* **39**, 491–501. (doi:10.3758/s13421-010-0039-y)
- 78 Gentner, T. Q., Fenn, K., Margoliash, D. & Nusbaum, H. C. 2010 Simple stimuli, simple strategies. *Proc. Natl Acad. Sci. USA* **107**, E65. (doi:10.1073/pnas.1000501107)
- 79 ten Cate, C., van Heijningen, C. & Zuidema, W. 2010 Reply to Gentner et al.: as simple as possible, but not simpler. *Proc. Natl Acad. Sci. USA* **107**, E66–E67. (doi:10.1073/pnas.1002174107)
- 80 Corballis, M. C. 2007 Recursion, language, and starlings. *Cogn. Sci.* **31**, 697–704. (doi:10.1080/15326900701399947)
- 81 Beckers, G. J. L., Bolhuis, J. J., Okanoya, K. & Berwick, R. C. 2012 Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* **23**, 139–145. (doi:10.1097/WNR.0b013e32834f1765)
- 82 Folia, V., Udd n, J., De Vries, M., Forkstam, C. & Petersson, K. M. 2010 Artificial language learning in adults and children. *Lang. Learn.* **60**, 188–220. (doi:10.1111/j.1467-9922.2010.00606.x)
- 83 Marcus, G. F., Fernandes, K. J. & Johnson, S. P. 2007 Infant rule learning facilitated by speech. *Psychol. Sci.* **18**, 387–391. (doi:10.1111/j.1467-9280.2007.01910.x)
- 84 Kroodsma, D. E. 1989 Suggested experimental designs for song playbacks. *Anim. Behav.* **37**, 600–609. (doi:10.1016/0003-3472(89)90039-0)
- 85 Kroodsma, D. E. 1990 Using appropriate experimental designs for intended hypotheses in ‘song’ playbacks, with examples for testing effects of song repertoire sizes. *Anim. Behav.* **40**, 1138–1150. (doi:10.1016/S0003-3472(05)80180-0)