Birdsong neurolinguistics: songbird context-free grammar claim is premature

Gabriël J.L. Beckers[a], Johan J. Bolhuis[b], Kazuo Okanoya[c,d] and Robert C. Berwick[e]

There are remarkable behavioral, neural, and genetic similarities between song learning in songbirds and speech acquisition in human infants. Previously, we have argued that this parallel cannot be extended to the level of sentence syntax. Although birdsong can indeed have a complex structure, it lacks the combinatorial complexity of human language syntax. Recently, this conclusion has been challenged by a report purporting to show that songbirds can learn so-called context-free syntactic rules and then use them to discriminate particular syllable patterns. Here, we demonstrate that the design of this study is inadequate to draw such a conclusion, and offer alternative explanations for the experimental results that do not require the acquisition and use of context-free grammar rules or a grammar of any kind, only the simpler hypothesis of acoustic similarity matching. We conclude that the evolution of vocal learning involves both neural homologies and behavioral convergence, and that human language reflects a unique cognitive capacity. *NeuroReport* 23:139–145 © 2012 Wolters Kluwer Health | Lippincott Williams & Wilkins.

Darwin noticed the parallels between language acquisition in human infants and song learning in songbirds [1]. Interestingly, vocal imitation appears to be absent in our closest relatives, the great apes, suggesting that there has been evolutionary convergence for this particular trait [2,3]. Since Darwin, it has become clear that in addition to behavior, there are also similar neural and genetic mechanisms underlying birdsong and speech that are shared with nonhuman primates [3,4]. After briefly reviewing these parallels, we will discuss the more contentious issue of putative linguistic similarities between birdsong and human language [5–8]. In particular, we critically evaluate a recent key study claiming to demonstrate the ability to acquire and use syntactic rules in Bengalese finches [9].

Similarities between song and speech

Behaviorally, there are at least three ways in which song learning in songbirds and speech acquisition in human infants are similar [2,3]. First, both human infants and songbirds acquire their speech and song repertoire, respectively, through a combination of predispositions and learning. Second, in both songbirds and humans, there is a sensitive period early in development during which auditory–vocal learning is optimal. Third, in both human infants and juvenile songbirds, there is a transitional phase at the start of vocal production, where the vocalizations are quite different from those of adult conspecifics. During this transitional phase, which is called 'babbling' in human infants and 'subsong' in songbirds, the young individual's vocalizations gradually come to resemble the adult form.

The songbird brain has discrete brain regions involved in song perception, production, and learning (see Fig. 1a and b). Essentially, birdsong involves three interconnected neural networks [3,4,10]. First, secondary auditory regions, including the caudomedial nidopallium (NCM) and caudomedial mesopallium, are involved in song perception and are important for the recognition of tutor song. Second, the song motor pathway (SMP) is involved in song production and certain aspects of song learning (Fig. 1b) [11]. The SMP is a posterior motor pathway connecting the HVC (acronym used as a proper name), the robust nucleus of the arcopallium (RA), and the tracheosyringeal portion of the nucleus hypoglossus (nXIIts). Third, the anterior forebrain pathway (AFP) is essential for sensori-motor learning and adult song plasticity [12]. The AFP is an anterior cortical–basal ganglia–thalamic loop that originates in HVC and passes through area X (part of the avian basal ganglia [12]), the thalamic nucleus dorsolateralis anterior, pars medialis, and the lateral magnocellular nucleus of the anterior nidopallium (LMAN), and eventually connects with the motor pathway at the robust nucleus of the arcopallium. Together, the SMP and AFP pathways are usually called the 'song system'.

In humans, traditionally, the neural substrate of motor representations of speech is considered to involve Broca's area in the inferior frontal cortex, whereas perception and memory of speech is thought to involve Wernicke's area
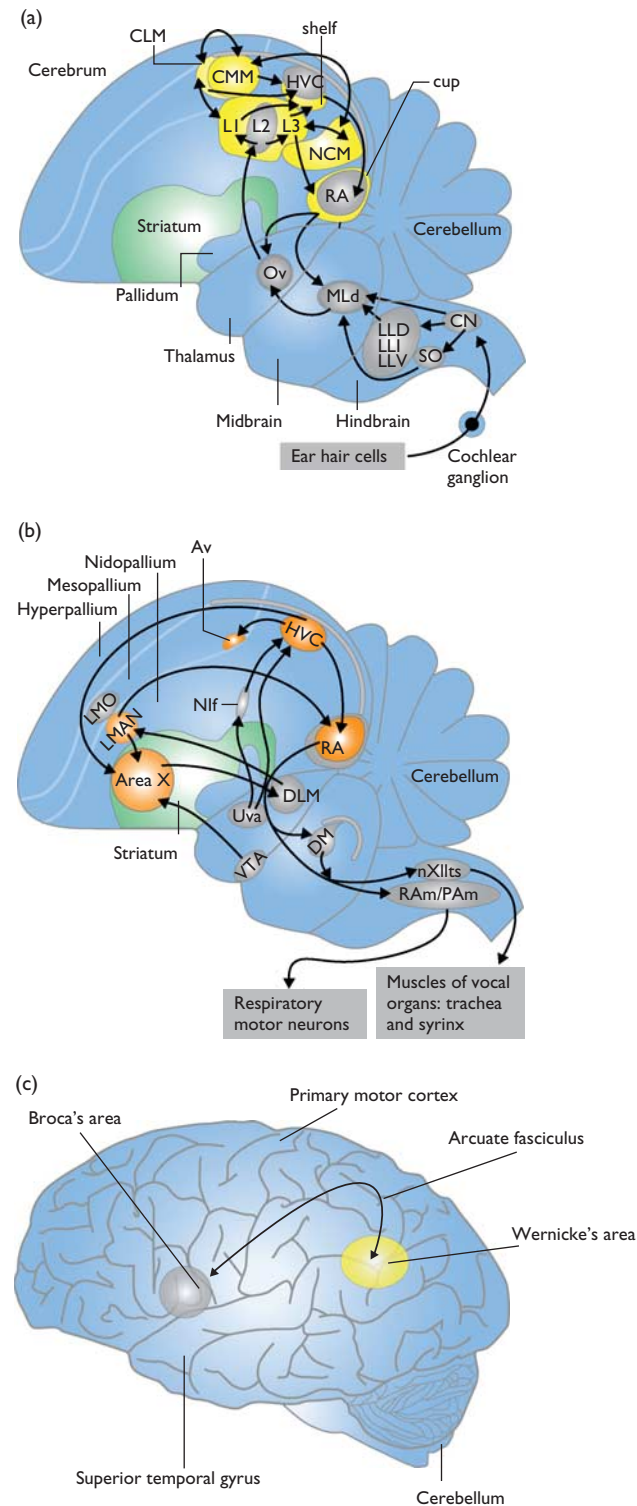
and surrounding regions in the superior temporal cortex. There are many analogies and homologies between the brains of birds and mammals, which have recently prompted a complete revision of the nomenclature of the avian brain [13]. Similarities in connectivity and function would suggest at least analogies between the human neocortex and the avian pallium (including the hyperpallium, mesopallium, nidopallium and arcopallium) [3,4]. Bolhuis and Gahr [4] have suggested that the projection regions of the field L complex – the NCM and the caudomedial mesopallium – may be analogous to the mammalian auditory association cortex. Doupe *et al.* [12] have argued that the AFP loop in the song system bears strong similarities in connectivity, neurochemistry, and neuron types to the mammalian basal ganglia, whereas both LMAN and HVC have been tentatively suggested to correspond functionally to Broca's area (see [3] for further discussion).

In addition to these neuroanatomical parallels, there is increasing evidence for a similar neural dissociation between auditory recognition and vocal production regions in the brains of songbirds and humans [3,4,14]. In the songbird, regions in the caudomedial pallium (including the NCM) contain the neural representation of tutor song memory that juveniles acquire [3,4,14,15], whereas nuclei in the song system are required for sensorimotor learning and song production [11].

A similar dissociation between brain regions involved in auditory perception and memory, on the one hand, and vocal production, on the other, exists in human speech. Human newborns show increased neural activity in the superior temporal lobe, but not in the inferior frontal cortex, in response to human speech [16]. A functional MRI study in 3-month-old infants (who are in the

Fig. 1

Schematic diagrams of composite views of parasagittal sections of the songbird brain and the human brain. (a) Diagram of a songbird brain giving approximate positions of nuclei and brain regions involved in auditory perception and memory. Yellow areas represent brain regions that show increased neuronal activation when the bird hears a song. (b) Diagram of a songbird brain giving approximate positions of nuclei and brain regions involved in vocal production and sensorimotor learning. The orange nuclei in the song system show increased neuronal activation when the bird is singing. (c) Schematic view of the left side of the human brain, with regions that are involved in speech and language (see text for details). Area X, area X of the striatum; Av, avalanche; CLM, caudolateral mesopallium; CN, cochlear nucleus; DLM, medial subdivision of the dorsolateral nucleus of the anterior thalamus; DM, dorsomedial subdivision of nucleus intercollicularis of the mesencephalon; HVC, a letter-based name; L1, L2, and L3 are subdivisions of field L; LLD, lateral lemniscus, dorsal nucleus; LLI, lateral lemniscus, intermediate nucleus; LLV, lateral lemniscus, ventral nucleus; LMAN, lateral magnocellular nucleus of the anterior nidopallium; LMO, lateral oval nucleus of the mesopallium; MLd, dorsal part of the lateral nucleus of the mesencephalon; NIf, interfacial nucleus of the nidopallium; nXIIts, tracheosyringeal portion of the nucleus hypoglossus (nucleus XII); Ov, nucleus ovoidalis; PAm, nucleus para-ambiguus medullaris; RA, robust nucleus of the arcopallium; RAm, nucleus retroambiguus medullaris; SO, superior olive; Uva, nucleus uvaeformis; VTA, ventral tegmental area. (Modified and reproduced with permission from Bolhuis *et al.* [3] and Moorman *et al.* [10], copyright 2010 Nature Publishing Group. All rights reserved.)

'cooing' stage of babbling in which syllables are not yet produced) showed activation in the superior temporal cortex (including Wernicke's area), as well as in Broca's area in response to hearing speech [17]. Six and 12-month-old infants exhibited increased activation (compared with baseline) in both Wernicke's and Broca's areas when exposed to speech sounds [16]. Together, these studies suggest that Wernicke's area is (part of) the neural substrate for speech perception in neonates and that Broca's area becomes active at a later stage, when infants start babbling [3].

A bird's eye view of language

In a recent paper, Abe and Watanabe [9] reported that Bengalese finches, *Lonchura striata* var. *domestica*, can discriminate vocal syllable strings according to spontaneously acquired grammatical rules, and that the brain region around LMAN is involved in such syntactic discrimination. Most notably, they found that these songbirds can spontaneously acquire and then use a center-embedding syntax, which has previously been argued to be unique to human language [18]. Abe and Watanabe's claim of this level of syntactic competence in birds, and their identification of a potential neural substrate, hold great promise for extending the songbird model system to include research into language-related neural mechanisms. However, here, we critically re-examine the experiment and the findings on which this claim is based, and conclude that this claim is premature.

To understand the importance of Abe and Watanabe's result, it is helpful to situate it within the standard framework of the formal analysis of language. Formally, a 'language' is some set of strings, or 'sentences', defined over an (fixed) alphabet, where the alphabet consists, for instance, of the distinct syllables in a birdsong or the distinct words in a human language. Thus, we might describe a particular birdsong language as consisting of 'sentences' with any number of 'warble' syllables, w, followed by an ending coda syllable, f. Such a birdsong

Fig. 2
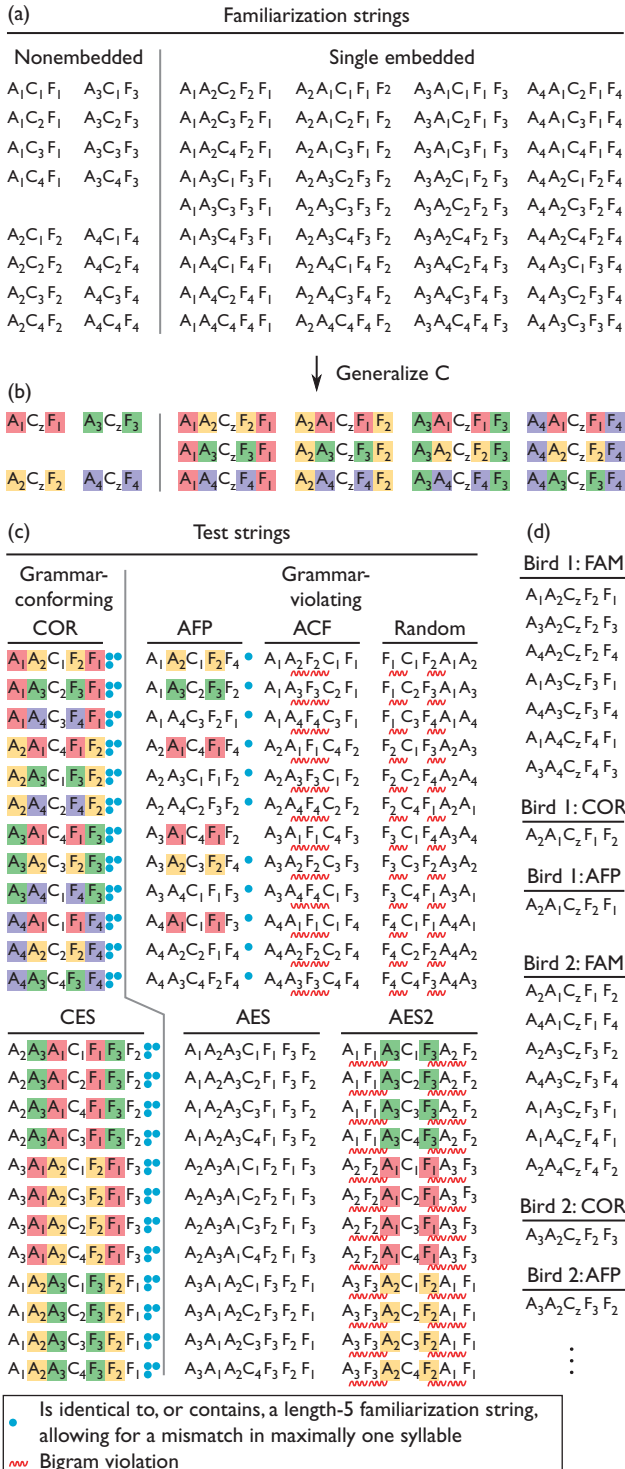
Song syllable strings used in Abe and Watanabe's [11] experiment to show that Bengalese finches spontaneously learn and used strictly context-free center-embedding grammar (the strings in A and C were kindly provided by these authors). (a) Familiarization strings; letters with specific subscripts, for example A_1, stand for specific song syllable sounds, and 3-g (nonembedded) and 5-g (single embedded) syllable sequences for artificial songs that the birds were familiarized with. All strings conform to context-free center-embedding grammar. (b) If the contents of the central C syllable are generalized, only 16 unique familiarization strings remain. Note that the four nonembedded ones all occur in the embedded ones. (c) Test strings that were used to probe for the birds' learning and use of grammar. There were two groups of grammar-conforming strings (COR and CES) and five groups of grammar-violating strings (AFP, ACF, random, AES, AES2). Note that all COR strings fully match with a length-5 C-generalized familiarization string, whereas CES contains such strings. Grammar-violating test strings lack this level of similarity. A blue dot indicates that a test string is identical to, or contains, a length-5 familiarization string, allowing for at most a mismatch in one syllable at any position (multiple dots indicate similarities to more than one familiarization string). (d) An outline of a test for the use of context-free center-embedding grammar that is not confounded by unbalanced phonetic resemblances. ACF, A–C–F sequence rule violations; AES, abnormally embedded strings; AES2, abnormally embedded strings and A–C–F sequence violations; AFP, A–F pairing rule violations; CES, correctly embedded strings; COR, correct strings.

language would contain an infinite number of 'sentences' or songs: *wf, wwf, wwwf*, and so forth. Formally, languages are generated by grammar, where grammar consists of a (finite) set of rules, and generation means applying the rules in some particular order, beginning with a special 'Start' rule, ultimately arriving at any one particular sentence. Here, we need to consider just two distinct types of grammars: (i) the 'finite-state' (more correctly, 'right-regular' or simply 'regular') grammars and (ii) the 'context-free' grammars. A long-established result [19] holds that the languages generated by the context-free grammars strictly contain those generated by the finite-state grammars. Informally, regular grammars generate languages defined by 'flat' or 'linear' dependencies. Thus, for example, regular grammars can enforce the constraint that all strings of syllables (a song) begin and end with exactly one 'warble' or that in a song every 'warble' is always followed by exactly one 'tweet', ending with a special syllable *f*, an adjacency restriction known as a 'bigram dependency'. Formally, (right)-regular grammars consist of a finite set of rules in the form $X \rightarrow wY$ or $X \rightarrow w$, along with a special starting rule $Start \rightarrow wX$, where *w* ranges over any word in the alphabet and *X* or *Y* ranges over some finite set of nonalphabet symbols or states. We may envision the generation of a sentence as beginning at grammar's 'Start' state, with the arrow indicating a transition to the state after the arrow if we apply the rule, along with the production of the first word in the sentence, *w*. Then, from the state *X*, we may apply the rule $X \rightarrow wY$ and move to the state *Y*, at the same time generating the succeeding word *w*, and so on, until we apply a rule in the form $X \rightarrow w$ that generates only an alphabet symbol. Alternatively, we can think of the rule as an instruction to replace the symbol on the left-hand side of the arrow with the symbols on the right-hand side of a rule. Then we begin with the start symbol, 'Start', and replace that symbol with the symbols *wX* on the right-hand side of the 'Start' rule. Next, *X* may be replaced again with *wX*, and so our string *wX* becomes *wwX*, and so on, until we replace *X* with just *w*. The sequence of words thus generated constitutes a full sentence, and the entire set of possible sentences derivable from the 'Start' symbol constitutes the language generated by the grammar. For example, a regular grammar that generates our fanciful birdsong language of any number of warbles, *w*, ending with just one *f* could consist of just three rules: $Start \rightarrow wX$, $X \rightarrow wX$, and $X \rightarrow fF$. All evidence so far suggests that the phonological syntax of natural birdsong can be generated by such regular grammars [8].

In contrast, context-free grammars admit a more general rule format where, along with rules in the form $X \rightarrow wY$ and $X \rightarrow w$, we may also have rules in the form $X \rightarrow wYw$ or $Y \rightarrow wXw$, that is, with alphabet symbols on both sides of a state symbol *X* or *Y*. This more general rule type can encode 'nested' or 'center-embedded' dependencies, because words can appear on both sides of a state symbol

X or *Y*. In this way, context-free grammars can require that two particular words, say *a* and *f*, appear 'paired' or 'nested' on both sides of an *X*, by a rule in the form $X \rightarrow aXf$. This is a 'center-embedded' dependency, because the relationship recorded in the rule between *a* and *f* contains an intervening, embedded state symbol *X*. By applying such a rule once, we obtain the form *aXf*, where the first *a* is paired with the first *f*. If we apply the rule again, we obtain the form *aaXff*, where the innermost *a* is paired with the innermost *f* by the second application of the $X \rightarrow aXf$ rule, and the outermost *a* is paired with the outermost *f* by the first application of that rule. We could continue such 'self-embedding' indefinitely until we apply some rule that reaches a final state as above, say, by the rule $X \rightarrow c$, obtaining a sentence in the form *aaa...acfff...f*, where there are an equal number of *a*'s and *f*'s, and, crucially, the innermost *a* is paired with the innermost *f* by the next to last application of the rule $X \rightarrow aXf$, the second to innermost *a* is paired with the second to innermost *f* by the third from the last application of the same rule, and so on. To highlight this dependency structure more clearly, it is helpful to denote the paired occurrences of *a*'s and *f*'s by subscripts, for example $a_1 a_2 a_3 c f_1 f_2 f_3$. Because the matching *a*'s and *f*'s may be extended indefinitely in such context-free grammars, the language thus generated is strictly context free because it cannot be generated by any regular grammar. Informally, this is so because in order to match an arbitrary a_i with its f_i mate, we would need to 'remember' an arbitrary number of such *a*'s on the left, before starting to pair them up with their proper *f*'s on the right, requiring more than a finite number of states.

Context-free grammar in birds and humans

The acquisition and use of context-free grammars (or their corresponding languages) has so far never been described in the natural communication system of any nonhuman species. Of course, it does not follow that animals must therefore be incapable of mastering such grammar. Indeed, it has already been claimed on the basis of a study using artificial song stimuli and an operant task that starlings, *Sturnus vulgaris*, are able to learn a context-free language following a center-embedding syntax [5]. However, this claim remains controversial as this earlier experiment did not clearly pinpoint that the birds' recognition abilities could be unambiguously ascribed to the processing of the center-embedded structure [20]. This is because the artificial language that the birds were trained on did not strictly follow the 'nested' pattern described above, and thus the birds could have solved their task simply by counting, an already-established competence for birds [21]. The stimuli used in the recent Abe and Watanabe experiment, by contrast, were constructed according to the full set of rules that describes a center-embedding syntax, and, remarkably, the Bengalese finches acquired this syntax

spontaneously, without any reinforcement, and very quickly, with only a small number of training instances.

Turning to human participants, it has been known since at least the work of Miller and Isard [22] that people have great difficulty parsing both naturally occurring self-embedded sentences as well as center-embedded sentences constructed in artificial language learning experiments [22,23]. Confirming this, it has also long been known that people restructure sentences so as to avoid producing complex center-embedded structures, as well as spontaneously using alternative strategies for solving tasks that would otherwise provide evidence for the processing of such structures. Indeed, some researchers [24] have gone as far as to claim that people do not process center-embedded sentences successfully at all, in the sense intended by Watanabe and Abe. Then too, attempts at probing the ability of nonhuman primates in such artificial language learning settings have failed to unambiguously reveal the ability to process center-embedded forms [23–26].

Taken together, these earlier findings make Abe and Watanabe's recent results [9] all the more remarkable. Given the importance of the topic, it would seem essential to review the details of the Abe and Watanabe experiment, to determine whether it indeed leads to the conclusion they draw.

Context-free grammar in songbirds?

In Abe and Watanabe's [9] crucial experiment (described in their fig. 3), birds were exposed for 60 min to playback of two sets of familiarization strings, three-syllable-long strings (nonembedded strings) and five-syllable-long strings (single-embedded strings), both of which can be generated by a center-embedding context-free grammar in the manner described in the previous section. (The authors kindly provided us with these strings, which we reproduce in Fig. 2a and c.) Subsequently, birds received test strings, which consisted of two groups: the first had strings with a correct syntax, conforming to that of the familiarization strings; the second had an incorrect syntax, violating the pattern of the familiarization strings in some way. Each group was further subdivided into categories that differed according to how the strings did or did not conform to the correct syntax of the familiarization examples. There were two categories with a correct syntax, COR and CES, and five categories of syntactic violations, AFP, ACF, random, AES, and AES2 (Fig. 2c shows how these strings resemble or differ from the familiarization strings). The behavioral measure used for testing response differences in the test strings was the normalized difference between the number of calls that the birds produced in a 5-min silent period immediately before testing and during the 5-min test period (shift in call counts). The birds showed no significant difference in call response from the syntactically correct COR and CES strings, but did show a significant decrease in calling activity in response to the syntactically incorrect AFP,

ACF, AES, and AES2. This response difference formed the basis for Abe and Watanabe's claim that the birds spontaneously learned and used the center-embedding syntax of the familiarization strings.

However, after examining the specific strings used in Abe and Watanabe's experiment, it is possible to provide an alternative, more parsimonious explanation for these experimental results, which does not depend on the use of center-embedding context-free grammar or in fact on the use of any grammar at all. This explanation is perhaps most easily seen by noticing that if one ignored the contents of the central syllable C, the 16 nonembedded syllable strings would be reduced to a set of only four unique strings; similarly, the 36 embedded strings would be reduced to a set of 12 unique strings (compare Fig. 2a and b). Comparing the test strings with the C-generalized familiarization strings, we find that all test strings with correct syntax (COR and CES) but none of the syntax-violating test strings (AFP, ACF, random, AES, and AES) are identical to, or fully contain, C-generalized length-5 familiarization strings (compare colored A–F pairs between Fig. 2b and c). Further, even length-3 C-generalized familiarization strings occur only rarely within syntax-violating strings, whereas all correct test strings contain them. In particular, note that a specific letter with a specific numerical subscript in Fig. 2, for example A_1, stands for a specific sound, not a phonetic or a semantic category, so that the COR (syntactically correct) test string $A_1A_2C_1F_2F_1$ for the most part is identical to the familiarization string $A_1A_2C_2F_2F_1$ at the level of the acoustic pressure wave form.

Given the clear imbalance between syntactically correct and incorrect test strings with respect to acoustic similarity to the familiarization strings, we see at least two possible strategies that may account for the statistical differences in the Bengalese finches' response behavior without the use of any grammar.

First, as outlined in Figs 2a and 2b, the birds could have generalized the C syllable and then used memorized representations of the generalized familiarization strings for recognition. Generalization of the C syllable can be motivated by the fact that this is the only syllable that greatly reduces the memory footprint of the set of familiarization strings. If so, COR and CES resemble the five-syllable length familiarization strings to a much greater degree than any of the syntax-violating types, because all COR and CES strings are either identical to or contain such a string, whereas this is not the case for any syntax-violating string. However, a match with a shorter, three-syllable length familiarization string, C-generalized or not, would have to prove insufficient for the birds to treat test strings as familiar, because both the incorrect AES2 and AFP string types contain such matches. It is not unreasonable to assume that the birds would consider a five-syllable length match as more suitable for

recognition than a three-syllable length match, because longer nonrepetitive syllable strings are less likely to show a chance resemblance to other random strings than shorter strings. In addition, the three-syllable length familiarization strings ($n = 16$) were played less often than the five-syllable length familiarization strings ($n = 36$); the birds may have memorized the familiarization strings without special treatment of C and then simply matched test strings for the presence or absence of the memorized patterns, ignoring, or paying relatively little attention to, a mismatch in only one syllable. We indicate in Fig. 2c (blue dots) how often single-embedded familiarization strings occur in the strings of the given test sets, allowing for a mismatch in at most one of the five syllables at any position (i.e. not only C). Every syntax-conforming test string (COR and CES) matches with exactly three different familiarization strings. Syntactically ill-formed test strings, in sharp contrast, for the most part do not match with any familiarization string, the exception being AFP strings, 11 of which match, but then only with one familiarization string, not three. This high-level acoustic similarity between syntax-conforming test strings, which is lacking in syntax-violating test strings, may explain why the birds responded more to syntax-conforming strings.

These two alternative strategies do not require the use of any grammar, but simply the ability to recognize and discriminate sound patterns on the basis of sensory memory representations of their acoustic content, allowing for the generalization of, or differences in, minor parts of the stimulus. However, it is well known that nonhuman animals, including birds, may be able to 'count' [21] as well as exhibit sensitivity to syllable transition probabilities at a bigram level, and therefore, the participants in Abe and Watanabe's experiment may have additionally used such low-level syntactic cues. Consequently, a third explanation for the results may be that the birds used a combination of both memory, in terms of frequently recurring, fixed syllable patterns, as well as bigram statistics. For example, they may have formed a strong memory representation of the innermost A–F pairs of the familiarization strings. There are only four of them and they occur in every familiarization string (Fig. 2b). All grammar-conforming test strings contain such a pair, but the same is only true in one of the five categories of grammar-violating strings, namely AES2. However, in AES2, every string contains two A–F and two F–A transitions, which never occur in any familiarization string, and which thus violates bigram rules in four out of six syllable transition possibilities (bigram violations are indicated by red squiggles in Fig. 2b). This explanation could be seen as more complex than the previous two because it requires the use of a certain sort of grammatical rule. However, syllable bigrams are the simplest kind of pattern that can be generated by regular grammar, and, as noted in the previous section, are nowhere near the level of complexity that can be generated by strictly context-free grammar of the sort advanced by Abe and Watanabe. Figure 2d shows an example of how confounding bigram and multi-syllable acoustic similarity cues could be avoided.

These alternative explanations, which require only cognitive capacities that birds are known to possess, cast considerable doubt on Abe and Watanabe's claim that the birds spontaneously learned and then used center-embedding context-free grammar to discriminate the test sounds. To demonstrate definitively that the birds used such grammar, the stimulus set would have to be balanced so as to be able to factor out the acoustic confound of such familiarization cues. Ideally, the stimulus set should not contain them at all, but at a minimum, it seems unsuitable to use test strings that match familiarization strings in at least four out of five syllables, both in terms of acoustic content and syllabic position.

Conclusion
Designing an experiment that unambiguously demonstrates the use of center-embedding syntax remains more difficult than simply looking *a posteriori* for alternative explanations, and we do not intend to question the value of Abe and Watanabe's study. Nevertheless, we maintain that the methodological shortcomings outlined above cast doubts on their conclusion, given that the alternatives are more parsimonious. We believe that any test for context-free grammar capabilities in nonhuman animals should attempt to exclude phonetic cues as much as possible, so that the subjects must attend to the sequence structure of sounds in order to succeed at the defined task. Hence, more stringent tests are pivotal in establishing whether songbirds can indeed spontaneously acquire the ability to learn and use hierarchical structure or context-free rules.

What are the consequences of our critical evaluation of the Abe and Watanabe study [9] for the evolution of brain and cognition? An evolutionary scenario emerges where three factors are important. First, there is increasing evidence of neural and genetic homology, where similar genes and brain regions are involved in auditory learning and vocal production, not only in songbirds and humans but also in apes and mice. Second, there is evolutionary convergence with regard to the mechanisms of auditory–vocal learning, which proceeds in essentially the same way in songbirds and human infants, but not in apes or mice. Third, our reanalysis shows that recent claims for strictly context-free syntactic abilities in songbirds are premature, and that there is no evidence to suggest that nonhuman animals possess the combinatorial complexity of human language. As a consequence, presently, there is no credible animal model for the study of the neural substrate of human language syntax.

References

1 Darwin C. *The descent of man and selection in relation to sex.* London: Murray; 1882.

2 Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 1999; **22**:567–631.

3 Bolhuis JJ, Okanoya K, Scharff C. Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Rev Neurosci* 2010; **11**:747–759.

4 Bolhuis JJ, Gahr M. Neural mechanisms of birdsong memory. *Nat Rev Neurosci* 2006; **7**:347–357.

5 Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. Recursive syntactic pattern learning by songbirds. *Nature* 2006; **440**:1204–1207.

6 Okanoya K. Language evolution and an emergent property. *Curr Opin Neurobiol* 2007; **17**:271–276.

7 Van Heijningen CAA, de Visser J, Zuidema W, ten Cate C. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc Natl Acad Sci USA* 2009; **106**:20538–20543.

8 Berwick RC, Okanoya K, Beckers GJL, Bolhuis JJ. Songs to syntax: the linguistics of birdsong. *Trends Cogn Sci* 2011; **15**:113–121.

9 Abe K, Watanabe D. Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat Neurosci* 2011; **14**:1067–1074.

10 Moorman S, Mello CV, Bolhuis JJ. From songs to synapses: molecular mechanisms of birdsong memory. *BioEssays* 2011; **33**:377–385.

11 Brainard MS, Doupe AJ. Auditory feedback in learning and maintenance of vocal behaviour. *Nat Rev Neurosci* 2000; **1**:31–40.

12 Doupe AJ, Perkel DJ, Reiner A, Stern EA. Birdbrains could teach basal ganglia research a new song. *Trends Neurosci* 2005; **28**:353–363.

13 Jarvis ED, Güntürkün O, Bruce L, Csillag A, Karten H, Kuenzel W, *et al.* Avian brains and a new understanding of vertebrate brain evolution. *Nat Rev Neurosci* 2005; **6**:151–159.

14 Gobes SMH, Bolhuis JJ. Bird song memory: a neural dissociation between song recognition and production. *Curr Biol* 2007; **17**:789–793.

15 Gobes SMH, Zandbergen MA, Bolhuis JJ. Memory in the making: localized brain activation related to song learning in young songbirds. *Proc R Soc Lond B* 2010; **277**:3343–3351.

16 Imada T, Zhang Y, Cheour M, Taulu S, Ahonen A, Kuhl PK. Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport* 2006; **17**:957–962.

17 Dehaene-Lambertz G, Hertz-Pannier L, Dubois J, Mériaux S, Roche A, Sigman M, *et al.* Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc Natl Acad Sci USA* 2006; **103**:14240–14245.

18 Hauser MD, Chomsky N, Fitch TW. The faculty of language: what is it, who has it, and how did it evolve? *Science* 2002; **298**:1569–1579.

19 Chomsky N. Three models for the description of language. *IRE Trans Info Theor* 1956; **2**:113–124.

20 Corballis MC. Recursion, language and starlings. *Cogn Sci* 2007; **31**:697–704.

21 Koehler O. The ability of birds to 'count'. *Bull Anim Behav* 1950; **9**:41–45.

22 Miller GA, Isard S. Free recall of self-embedded English sentences. *Inform Control* 1964; **7**:292–303.

23 Friederici AD, Bahlmann J, Friedrich R, Makuuchi M. The neural basis of recursion and complex syntactic hierarchy. *Biolinguistics* 2011; **5**:87–104.

24 Perruchet P, Rey A. Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychon Bull Rev* 2005; **12**:307–313.

25 Miller G, Chomsky N. Introduction to the formal analysis of natural languages. In: Luce R, Bush R, Galanter E, editors. *Handbook of mathematical psychology.* New York: Wiley; 1963. pp. 269–323.

26 Langedoen T. Finite-state parsing and the status of readjustment rules in the lexicon. *Ling Inq* 1975; **6**:533–554.