PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Review

Cite this article: Rohrmeier M, Zuidema W, Wiggins GA, Scharff C. 2015 Principles of structure building in music, language and animal song. *Phil. Trans. R. Soc. B* **370**: 20140097.
http://dx.doi.org/10.1098/rstb.2014.0097

One contribution of 12 to a theme issue 'Biology, cognition and origins of musicality'.

Subject Areas:
cognition, computational biology, neuroscience

Keywords:
music, animal vocalization, language, Chomsky hierarchy, comparative perspective, computational modelling

Author for correspondence:
Martin Rohrmeier
e-mail: martin.rohrmeier@tu-dresden.de

Principles of structure building in music, language and animal song

Martin Rohrmeier[1], Willem Zuidema[2], Geraint A. Wiggins[3] and Constance Scharff[4]

[1]Institut für Kunst- und Musikwissenschaft, Technische Universität Dresden, August-Bebel-Straße 20, 01219 Dresden, Germany
[2]ILLC, University of Amsterdam, PO Box 94242, 1090 CE Amsterdam, The Netherlands
[3]School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road, London E1 4FZ, UK
[4]Animal Behavior, Freie Universität Berlin, Takustraße 6, 14195 Berlin, Germany

Human language, music and a variety of animal vocalizations constitute ways of sonic communication that exhibit remarkable structural complexity. While the complexities of language and possible parallels in animal communication have been discussed intensively, reflections on the complexity of music and animal song, and their comparisons, are underrepresented. In some ways, music and animal songs are more comparable to each other than to language as propositional semantics cannot be used as indicator of communicative success or wellformedness, and notions of grammaticality are less easily defined. This review brings together accounts of the principles of structure building in music and animal song. It relates them to corresponding models in formal language theory, the extended Chomsky hierarchy (CH), and their probabilistic counterparts. We further discuss common misunderstandings and shortcomings concerning the CH and suggest ways to move beyond. We discuss language, music and animal song in the context of their function and motivation and further integrate problems and issues that are less commonly addressed in the context of language, including continuous event spaces, features of sound and timbre, representation of temporality and interactions of multiple parallel feature streams. We discuss these aspects in the light of recent theoretical, cognitive, neuroscientific and modelling research in the domains of music, language and animal song.

1. Introduction

Human language, music and the complex vocal sequences of animal songs constitute ways of sonic communication that have evolved a remarkable degree of structural complexity. Recent discussions have focused on comparing the structure of human language to that of learned animal songs, focusing not only particularly on songbirds, but also on whales and bats [1–4]. Such comparisons addressed aspects of phonology [5,6] and syntax [7–10], aiming to distinguish features that may characterize species-specific principles of structure building and reveal whether there might also be universal principles underlying the sequential organization of complex communication sounds. One debate concerns the role of 'recursion' as a core mechanism of the language faculty in the narrow sense [11] that is unique both to humans and to language.

In this review, we argue that although language and recursion are important topics for comparative studies, comparisons between human music and learned songs in animals deserve more attention than they are currently receiving (see also [12–14]). Structurally and functionally, music, language and animal songs not only share certain aspects but also have important differences. A three-way comparison between language, music and animal songs therefore has the potential to benefit research in all three domains, by highlighting shared and unique mechanisms as well as hidden assumptions in current research paradigms.

THE ROYAL SOCIETY PUBLISHING

We present an approach to performing such comparisons focusing on the structural organization of music, language and animal songs as well as the underlying function, motivation and context. We focus on models of structure and structure building in these domains and discuss their relations to empirical findings. Our starting point is the influential work of Shannon and Chomsky from the 1940s and 1950s. We discuss issues concerning building blocks, Shannon's *n*-gram models and the Chomsky hierarchy (CH) as well as their more recent extensions. Subsequently, we discuss limitations of both frameworks in relation to empirical observations from the biological and cognitive sciences and suggest ways for future research to move beyond these frameworks.

2. Building blocks and sequential structure

Before discussing models of structure building and sequence generation, we must first consider what the building blocks are from which sequences, be it in language, music and animal vocalizations, are built. One of the classic universal 'design features' of natural languages is 'duality of patterning' [15], which refers to the fact that all languages show evidence of at least two combinatorial systems: one where meaningless units of sounds are combined into words and morphemes, and one where those meaningful words and morphemes are further combined into words, phrases, sentences and discourse. While many of the details are debated, the assumption that words are building blocks of sentences is uncontroversial [16]. Like language, music and animal songs combine units of sound into larger units in a hierarchical way, but the comparability of the building blocks and the nature of the hierarchies of language, music and animal songs is not at all straightforward. In particular, there is no clear analogue of a 'word' in music or animal songs [17,18].

From a comparative perspective, some principles of structure building may trace back to evolutionary ancient cognitive principles, whereas some structures are coined by cultural effects (such as the impact of a formal teaching tradition or notation systems on musical or linguistic structure). One recent cross-cultural review proposes a list of statistical universals of musical structure; specifically, these include the use of discrete pitches, octave equivalence, transposability, scales that commonly have seven or fewer pitches in unequal steps, the use of melody, pitch combination rules and motivic patterns as well as the use of timing, duration and beat [19] (see also [20]). The pitch continuum is discretized (based on culturally established musical scales), and so is the timing continuum by a process of beat induction (which establishes a kind of grid overlaying a musical sequence [21]). There is further some work on cross-cultural principles of melodic structure that originated from Narmour's work [22,23]. Although much more cross-cultural research is necessary [24], in the remainder of this review, we take as established that there is evidence for a number of structure building operations at work in music, which we would like to account for with formal models. These include repetition and variation [25], element-to-element implication (e.g. note–note, chord–chord) [22,26], hierarchical organization and tree structure as well as nested dependencies and insertions [27,28].

One common thread in this work is that at any point in a musical sequence listeners are computing expectations about how the sequence will continue, both regarding timing details and regarding classes of pitches or other building blocks [26,29]. Composers can play with these expectations: meet expectations, violate them or even put them on hold. Traditionally, much of the discussion of structure in music has focused on Western classical music and has built on building blocks of melody, voice-leading [30–32], outer voices [33], harmony [34–36], combinations of harmony and voice-leading [33,37,38], or complex feature combinations derived from monophonic melody [39,40] and harmony [41,42]. Although observations about Western music are clearly relevant to understanding human musical abilities, some of the complexity we find in Western music might for instance be influenced by (culture-specific) factors such as notation systems and aesthetic principles that may not necessarily generalize with respect to cognitive processes and cross-cultural principles.

In the domain of animal vocalization, there is plenty of evidence of non-trivial structure building, but the choice of building blocks for analysis is far from clear. We focus the discussion in this review primarily on learned songs of animals, which provide the most interesting comparison with language and music, as structurally particularly complex and largely learned traits (and we therefore do not say much about animal *calls*, which depend less or not on learning, and are perhaps more usefully compared with other, innate, human utterances such as cries, sighs and laughter). Animal songs occur primarily in a territorial or courtship setting (fighting and flirting) and are produced in the majority of bird species by both sexes [43].

Songs usually consist of longer strings of temporally ordered, recurring, acoustically distinct elements. In contrast, calls occur in different contexts (e.g. contact, begging, alarm) and tend to be temporally and spectrally more variable [44]. Animal songs contain hierarchically structured elements (often called 'notes') that can be combined into 'syllables' or 'note complexes' which in turn can constitute 'motifs' or 'song types'. Note that in contrast to language, repetition with little or no variation of smaller or larger units of building blocks, somewhat similar to music [25], are a typical feature of song in many bird species (figure 1).

Animal songs are the product of natural and sexual selection and continue to be constrained by these forces. Finding similar structure building principles in different bird species would argue for common neurocognitive foundations for birdsong that might have parallels with those underlying human musicality [45]. In addition, research on female preference in different bird species has already revealed which elements of courtship song are particularly relevant and attractive, and those features consequently change little during the course of evolution, whereas other song features are more free to drift. Conversely, there is clear evidence from many songbird species for female preference of larger rather than smaller repertoire sizes, leading potentially to more combinatorial possibilities in sequencing and thus driving the evolution of syntactic complexity.[1]

Some arguments concerning building blocks have been made based on neuroscience and motor control evidence rather than by analysis of sequences. Patel and co-workers have argued that motor constraints are similar in human and bird song and focused on the note, for example a bioacoustic gesture surrounded by silent gaps needed to inhale as the basic unit of comparison [46]. This, however, differs between bird species. Others have argued that the basic units in birdsong are smaller gestures coded by different neural

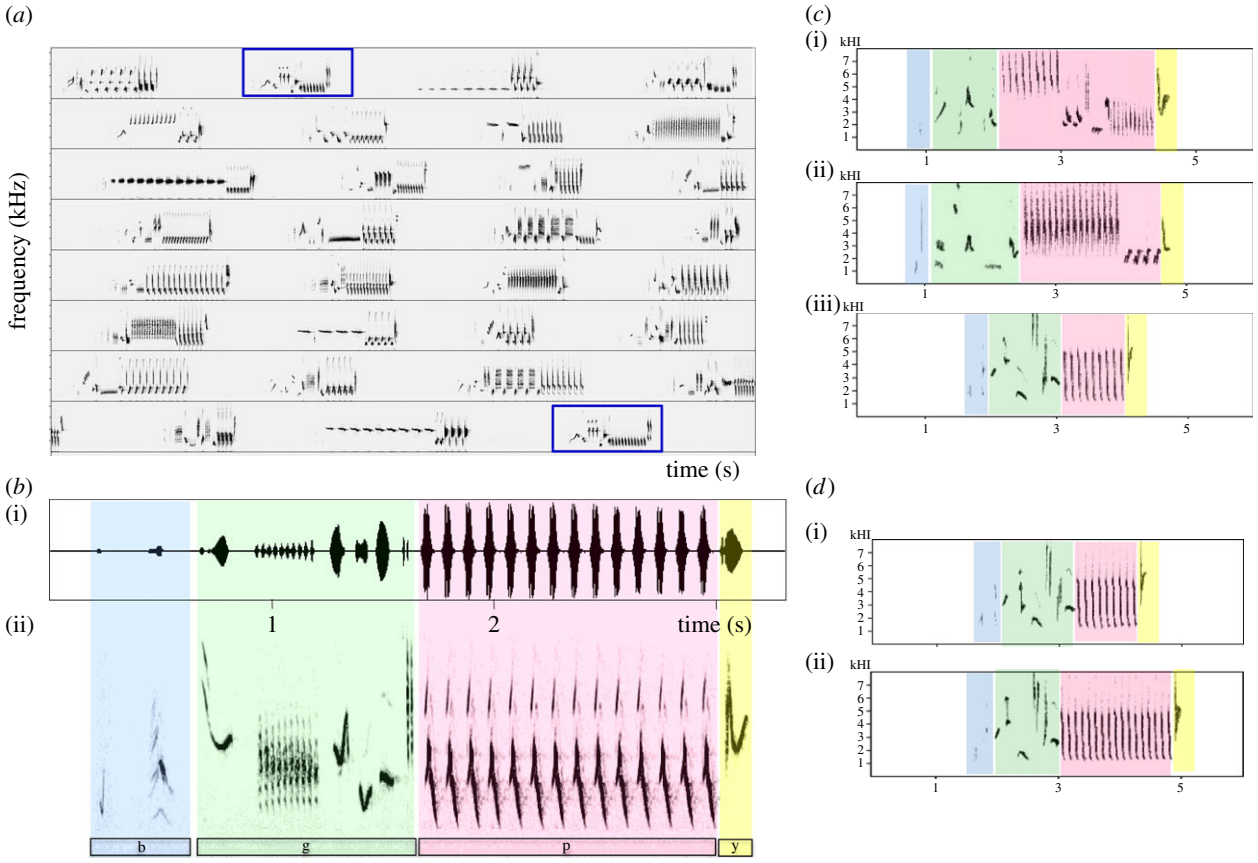

Figure 1. Hierarchical organization of nightingale song. Panel (*a*) depicts a spectrogram of *ca* 2 minutes of continuous nocturnal singing of a male nightingale. Shown are 30 sequentially delivered unique songs. The 31st song is the same song type as the second one (both framed). The average repertoire of a male contains about 150 unique song types, which can be delivered in variable but non-random order for hours continuously. Panel (*b*) illustrates the structural components of one song type. Individual sound elements are sung at different loudness (amplitude envelope in (i)) and are acoustically distinct in the frequency range, modulation, emphasis and temporal characteristics (spectrogram in (ii)). Panel (*c*) illustrates the structural similarities in three different song types (i,ii,iii). Song types begin usually with one or more very softly sung elements (blue, b), followed by a sequence of distinct individual elements of variable loudness (green, g). All song types contain one or more sequences of loud note repetitions (pink, p) and are usually ended by a single, acoustically distinct element (yellow, y). Panel (*d*) illustrates that the same song type (i,ii) can vary in the number of element repetitions in the repeated section (pink). Spectrograms courtesy of Henrike Hultsch. (Online version in colour.)

ensembles and not necessarily separated by silent gaps [47]. Further larger units of sound sequences (often referred to as 'chunks' or 'song types' or 'motifs') are used as building blocks for recombination. This is suggested by non-random association of particular sequences in various bird species, to which we return below.

In summary, the issues and challenges concerning the choice of building blocks are fundamental for the forms of sequential model and model representation we discuss below. However, the present state of the art offers no final word that may settle the differences between proposals for building blocks and the relations between different choices. It remains a matter of future research to explore the extent to which different types of music and animal vocalizations may require different choices of building blocks and to perform an empirical comparison of different approaches: although it seems natural to want to settle on the question of what the building blocks are before investigating how these building blocks are combined into sequences, it might be necessary to study both issues at the same time.

3. Shannon's *n*-grams

Shannon [48] introduced, in the slipstream of his major work on information theory, *n*-grams as a simple model of sequential structure in language. *n*-Grams define the probability of generating the next symbol in a sequence in terms of the previous $(n-1)$ symbols generated. When $n=2$, *n*-grams become 'bigrams', which simply model transitional probabilities: the probability of generating the next word only depends on what the current word is. In this respect, *n*-gram models are equivalent to $(n-1)$th-order Markov models over the same alphabet. Shannon demonstrated with word counts from corpora that the higher n is, the better one can predict the next word (given a corpus that is sufficiently large). This insight still forms a crucial component in many engineering applications in speech recognition and machine translation [49].

n-Gram models (often simple bigrams) have also been frequently applied to bird song [50–54] and music [55,56] (see below for further details). For many bird species, bigrams in fact seem to give a very adequate description of the sequential structure. Chatfield & Lemon [51] studied the song of the cardinal, and reported that a 3-gram (trigram) model modelled song data only marginally better than a bigram model (measured by the likelihood of the data under each of these models). More recent work with birds raised with artificially constructed songs indicates that transitional probabilities between adjacent elements are the most important factor in the organization of the songs also in zebra finches and Bengalese finches [57], although there are also many examples of bird song requiring richer models as we discuss below [8,58–60].

In music research, numerous variants of *n*-gram models have been used predominantly in the context of modelling

predictive processing [29]. The perception of tonality and key has been argued to be governed by pitch distributions that in fact correspond with unigram models [61,62]. Narmour [22] proposed a number of principles that govern melodic structure across cultures, which later work showed can be simplified considerably and put in the form of a handcrafted bigram model over melodic intervals [63–66]. In the domain of harmony, Piston's [67] table of common root progressions and Rameau's [68] theory (of the *basse fondamentale*) may be argued to have the structure of a first-order Markov model (a bigram model) of the root notes of chords [69,70].

All of these theoretical approaches constitute variants of Markov/*n*-gram models that could be encompassed by overarching generalized Markov models that learn their parameters from a corpus [29,71,72]. In fact, Markov modelling of music has been carried out early (as early as we had computers in scientific use; cf. [55,56]). In particular, general *n*-th order Markov models have been proposed for melody and harmony [39–42,73–75] and employed for segmentation and boundary entropy detection [76]. In practical, ecological contexts, it is a common finding that *n*-gram models with large values of the context length *n* result in suboptimal models owing to sparsity issues or overfitting. In analogy with the findings in bird song research, several music modelling studies find trigrams optimal with respect to modelling melodic structure [77] or harmonic structure [42].

The musical surface, however, is more complex than ordinary language, or potentially animal song, because of the interaction with metrical structure, which means that surface symbols, such as notes or chords, do not all have the same salience when forming a sequence. This is demonstrated by a study of harmonic structure using a corpus of seventeenth century dance music [73]: an *n*-gram model taking into account three-beat metrical structure (and representing each beat by one symbol) seems to favour 4-grams. This is probably because the first beat of a bar is more musically salient than the other two, in harmonic terms, and a 4-gram is able to directly capture at least some of this importance, in this representation, where a 3-gram necessarily cannot.

4. The classical Chomsky hierarchy

Shannon's *n*-grams are simple and useful descriptions of some aspects of local sequential structure in animal communication, music and language and have also been discussed as simple cognitive models. But what are their limitations? In theoretical linguistics, *n*-grams, no matter how large their *n*, were famously dismissed as useful models of syntactic structure in language in the foundational work of Noam Chomsky from the mid-1950s [78]. In his work, Chomsky first argued against incorporating probabilities into language models; in his view, the core issues for linguists concern the symbolic, syntactic structure of language. Chomsky proposed an idealization of language where a natural language such as English or Hebrew is conceived of as a (potentially infinite) set of sentences, and a sentence is simply a sequence of words (or morphemes).

The CH concerns different classes of formal languages that generate such sets of sequences. In the classical formulation, it distinguishes four classes: regular languages, context-free languages, context-sensitive languages and recursively enumerable languages. Each class contains an infinite number of sets, and there are subset relations between the classes: every regular language is also context-free, every context-free language is also context-sensitive and every context-sensitive language is recursively enumerable. (*n*-Grams, when probabilities are stripped off, correspond to a true subset of the regular languages—see below.)

Table 1. Example sentence, the corresponding context-free grammar and the derivation showing centre-embedding.

example sentence	context-free rules
'either language came first or music came first'	(1a) S → either S or S
	(1b) S → NP VP
this sentence can be derived from the start symbol S by subsequently applying rules 1a,1b,2a,3,4,5,1b,2b,3,4,5. Note that rule 1a *center-embeds* a phrase of category S within a phrase of category S, which is beyond the power of finite-state automata	(2a) NP → language
	(2b) NP → music
	(3) VP → V ADV
	(4) V → came
	(5) ADV → first
'either language or music came first'	
to derive this second sentence (with a so-called *ellipsis*), whilst maintaining the subject relation between 'language' and 'came first', linguists have proposed formalisms even richer than context-free grammars	

For cognitive science, the relevance of the hierarchy comes from the fact that the four classes can be defined by the kinds of rules that generate structures as well as by the kind of computations needed to parse the sets of sequences in the class (the corresponding formal automaton). Informally, regular languages are the types of sets of sequences that can be characterized by a 'flowchart' description (finite-state automaton). Crucially, when generating or parsing the next word in a sentence of a regular language, we need only to know where we currently are on the flowchart, not how we got there.

In contrast, at all higher levels of the CH, some sort of memory is needed by the corresponding formal automaton that recognizes the language. The next level up in the classical CH are context-free languages, generated/recognized by context-free grammars (CFGs, equivalent to so-called 'push-down automata'). CFGs consist of (context-free) rewrite rules that specify which symbols (representing a category of words or other building blocks, or categories of phrases) can be rewritten to which list of symbols. Chomsky observed that natural language syntax allows for nesting of clauses (centre-embedding), and argued that the finite-state automata are inadequate to account for such phenomena. In contrast, CFGs can express multiple forms of nesting as well as forms of counting elements in a sequence. An example of such nesting, and a CFG that can describe it, is given in table 1. Table 1 further shows a linguistic example that requires even higher complexity.

The success of the CH in linguistics and computer science and Chomsky's negative demonstration that natural language syntax is beyond the power of finite-state automata has influenced many researchers to examine the formal structures

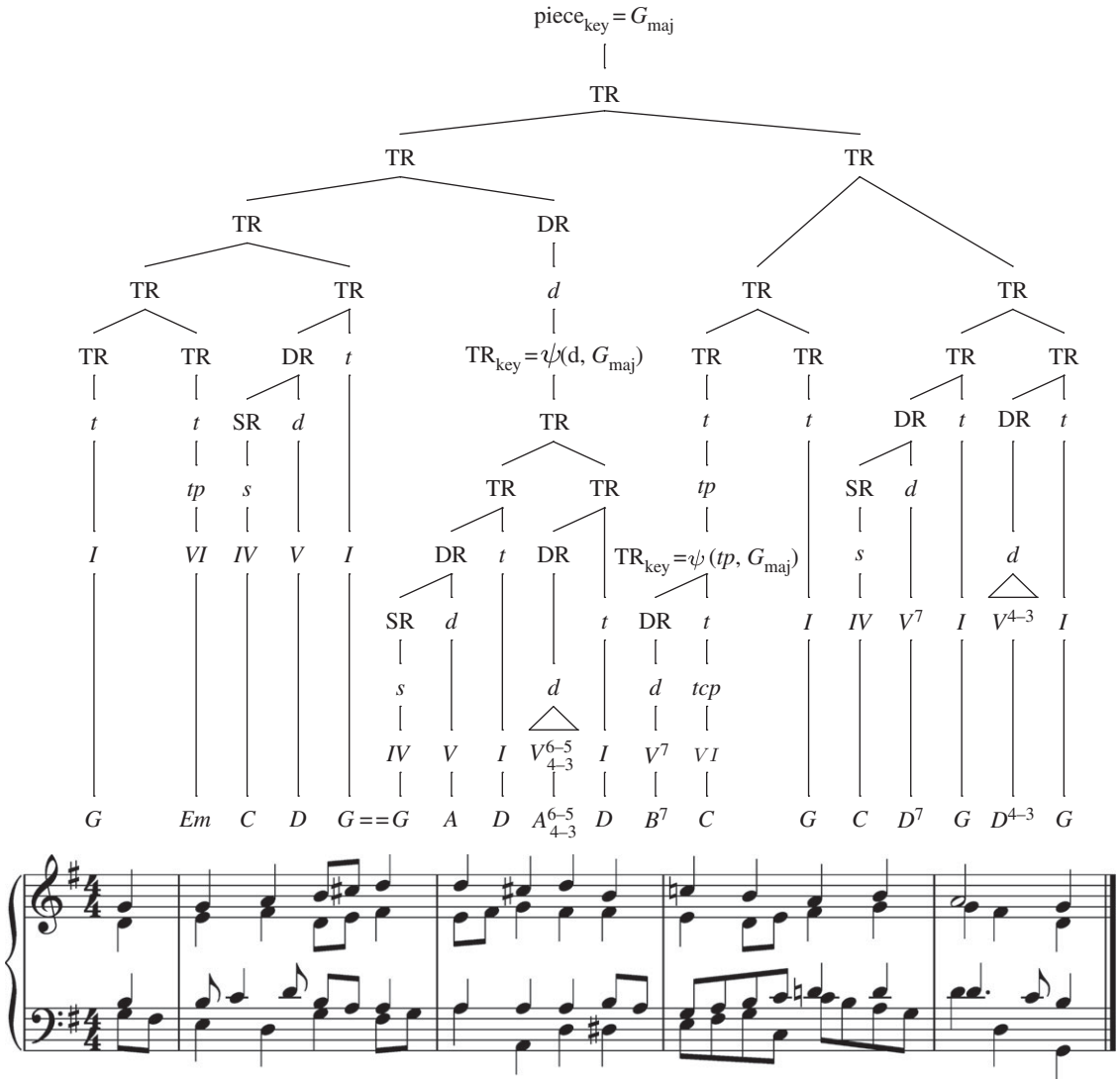

Figure 2. Analysis of Bachs chorale 'Ermuntre Dich, mein schwacher Geist' according to the GSM proposed by Rohrmeier [36]. The analysis illustrates hierarchical organization of tonal harmony in terms of piece (piece), phrases (*P*), functional regions (TR, DR, SR), scale-degree (roman numerals) and surface representations (chord symbols). The analysis further exhibits an instance of recursive centre-embedding in the context of modulation in tonal harmony. The transitions involving $TR_{key=\psi(x,ykey)}$ denote a change of key such that a new tonic region (TR) is instantiated from an overarching tonal context of the tonal function *x* in the key *ykey*.

underlying animal song and music (though there is no comprehensive comparison of music-theoretical approaches in terms of the CH yet). The advantages of CFGs to express hierarchical structures, categories of events and, particularly, recursive insertion/embedding have lent themselves to a number of theoretical approaches that characterized Western tonal music [22,34–38,79–84]. Schenker's [85] seminal work constitutes the foundation of hierarchical approaches to music. However, owing to its largely informal nature, it is not clear whether its complexity is adequately expressed by a CFG or whether it requires a more complex mechanism [86] (see also the related discussion in [87]).

The approaches by Rohrmeier [35,36] suggest a formal argument that musical modulation (change of key and centre of tonal reference) constitutes an instance of recursive context-free embedding of a new diatonic space into an overarching one (somewhat analogous to a relative clause in language), a point considered in informal terms already by Hofstadter [88]. Figure 2 shows an example of a syntactic analysis of the harmonic structure of a Bach chorale that illustrates an instance of recursive centre-embedding in the context of modulation. Given the absence of communication through propositional semantics in music (see below), the occurrence of nested context-free structure might be explained by patterns of implication and prolongation and associated features of tension: given that an event may be prolonged (extended through another event; an idea originating from Schenker [85]), and events may be prepared/implied by other events, the possibility of multiple and recursive preparations, combined with the combinatorial play of recursion and the option to employ an event as new tonal centre, established context-free complexity as well as a straight link to a motivation through complex patterns of musical tension [89,90]. Another potential explanation for the occurrence of complex hierarchical, recursive structure in Western tonal music may be found in the notation system and the tradition of formal teaching and writing—a factor that may even be relevant for complexity differences in written and spoken languages in communities that may differ with respect to their formal education [91].

There are also analytical findings that suggest that principles of hierarchical organization may be found in classical North Indian music [27] that is based on a tradition of extensive oral teaching. However, more cross-cultural research on other cultures and structure in more informal and improvised music is required before more detailed

conclusions may be drawn concerning structural complexity and cross-cultural comparisons.

There is currently little evidence that non-human structures or communicative abilities (either in production or in reception) exceed finite-state complexity. One structure that constitutes a core instance of context-free structures is A^nB^n (in which there is the same number of As and Bs; or $A_iA_jA_k \ldots B_kB_jB_i$ in which the string features pairs (A_x, B_x) in a retrograde structure). Claims have been made—and been refuted—that songbirds are able to learn such instances of context-free-structures (see [92,93]; and respective responses [94–96]). Hence, further targeted research with respect to transfinite-stateness of animal song is required to shed light on this question.[2] By contrast, there is a number of studies arguing for implicit acquisition of context-free structure (and even (mildly) context-sensitive structure) in humans in abstract stimulus materials from language and music [97–103].

5. Limitations of the Chomsky hierarchy

Several extensions to the CH have been proposed. We introduce these by first considering one common problem in the literature, trying to decide empirically where to place language, music and animal songs on the CH. This problem is that a number of related but different issues are often conflated:

(1) the inadequacy of (plain) n-gram models (but not necessarily of finite-state automata) for modelling music or animal songs;
(2) the presence of long-distance dependencies (and the ability of animal or human subjects to detect them); and
(3) the ability of subjects to process context-free languages.

Why issue (1) differs from (2) and (3) can be understood by considering the extended CH that introduces several different levels. Below the level of regular languages, ongoing research established the so-called subregular hierarchy that includes, among others, the class of strictly locally testable languages (SLTL) [104]. SLTLs constitute the non-probabilistic counterpart of n-gram models. SLTLs/n-grams do not assume underlying unobservable (hidden) structure and model the string language based on surface fragments alone. Regular languages and Markov models are therefore not equivalent, and showing the inadequacy of n-grams or SLTLs is hence not sufficient by itself to prove the need to move beyond finite-state or to the context-free level.

In reverse, the hierarchical nature of the extended CH (and methodological problems in dealing with real-world data that we outline below) creates potential issues for some arguments that the formalism producing a set of sequences is constituted by a lower class: the mere fact that a lower complexity model of structure could be built does not constitute a valid form of argument or proves that the system in question is best modelled by this type of complexity. Particularly, the fact that Markov models may be easily computed and used to describe some statistical features of corpora of music [26,105–107] does crucially not imply or even underpin an argument that a Markov model is the best model (in terms of strong generative power, compression or model comparison; see §§6,7). A Markov model may be computed from sequences generated from *any* deep structure and the corresponding models result in being oblivious to any other than local structure (e.g. forms of embedding or non-local dependencies).

With respect to animal songs and animal pattern recognition involving sequences of elementary units, it seems that n-gram models suffice for many species, but richer (but still finite-state) models are needed to characterize the songs of Bengalese finches [59,60], blackbirds [8,108] and other birds singing complex songs [109] (see [58] for a review). In particular, ten Cate *et al*. [8] argue that there are frequently observed phenomena in songs of songbird species such as blackbirds that require the power of hidden Markov models (probabilistic counterparts of regular grammars, see below); these phenomena include the optionality of elements and constraints on the number of repetitions in blackbird song.

Also in classes higher on the CH, more fine-grained distinctions have been introduced. In particular, a number of linguistic formalisms (e.g. tree-adjoining languages, combinatorial categorial grammar [110,111]) have been proposed that have been collectively referred to as mildly context-sensitive [112]. These grammars define classes of languages that are subsets of context-sensitive languages and supersets of the context-free languages. They share the expressive power to be able to express features such as cross-serial dependencies that human languages possess [113] and may be efficiently processable and learnable. We are not aware of any formally founded claims about mild context sensitivity in the domain of music or animal songs. Having introduced the extended CH, figure 3 provides a general overview of the locations of main results of structure building in language, music and animal song we discussed in the framework of the extended CH.

However, even when considering its extensions and despite its frequent use in recent cognitive debates, the CH may not be suited for providing a good class of cognitive or structural models that capture frequent structures in language, music and animal songs. One aspect stems from the fact that the CH is by its definition fundamentally tied to rewrite rules and the structures that different types of rewrite rules constrained by different restrictions may express. One well-known issue—and an aspect that the notion of mild context-sensitivity addresses—concerns the fact that repetition, repetition under a modification (such as musical transposition) and cross-serial dependencies constitute types of structures that require quite complex rewrite rules (see also the example of context-sensitive rewrite rules expressing cross-serial dependencies in reference [114]). In contrast, such phenomena are frequent forms of form-building in music [25] and animal song. This mismatch between the simplicity of repetitive structures and the high CH class it is mapped onto might be one of many motivations to move beyond its confines.

6. Moving towards different types of models

The CH has been extensively used in recent cognitive debates on human and animal cognitive capacities, discussing the complexity of theories and processes in various domains and in characterizing different types of structures that may be learnable in artificial grammar learning and implicit learning literatures [115]. However, discussions in many of these approaches relating to the CH resulted in complex confusions concerning the distinctions between the (formal) class of a language, the formal automata producing/accepting formal languages and the complexity involved in techniques to learn such structures

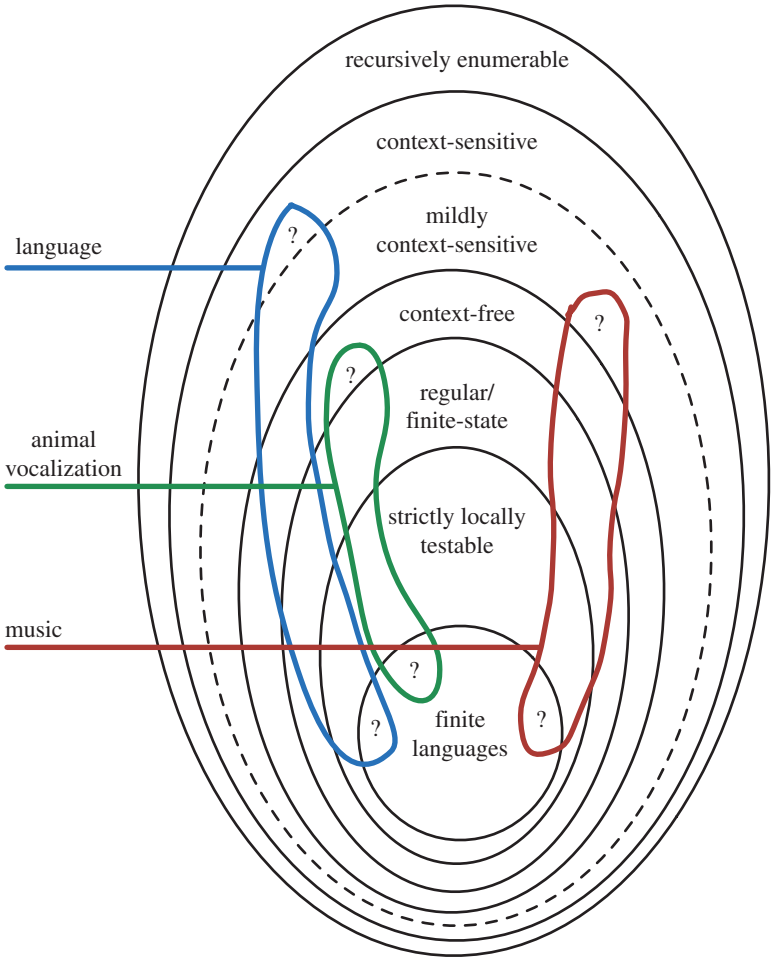

Figure 3. A Venn diagram of the Chomsky hierarchy of formal languages with three extensions annotated with a comparison of the hypothesized classifications of human languages, (human) music and animal vocalization. The areas marked with 'question' signs indicate that further research is required to settle examples for the respective class of complexity in these domains. (Online version in colour.)

from examples. It is important to distinguish here that while the formal automata that accept (or generate) a class of formal language in the CH may be comparably easy, the inference procedures to learn such structures from examples (e.g. to infer CFGs) are highly complex and are likely to be underestimated in artificial grammar learning research.[3]

Importantly, the CH concerns a theoretical construct that organizes types of structures according to different forms of rewrite rules and has, being a theory of formal languages in conjunction with idealized formal automata, little immediate connection with cognitive motivations or constraints (such as limited memory). The fact that it defines a set of languages that happen to be organized in mutual superset relations and are well explored in terms of formal automata that produce them does not motivate its reification in terms of mental processes, cognitive constraints or neural correlates. Although the CH has been inspired research in terms of a framework that allowed for the comparison of different models and formal negative arguments against the plausibility of certain formal languages or corresponding computational mechanisms, it does not constitute an inescapable *a priori* point of reference for all kinds of models of structure building or processing. Such forms of formal comparison and proofs should inspire future modelling endeavours, yet better forms of structural or cognitive models may involve distinctions orthogonal to the CH and may rather be designed and evaluated in the light of modelling data and its inherent structure as well as possible.

What are some different aspects that new models of structure building and corresponding cognitive models should take into account? In order to model the complexity of ecological real-word structures, they should be able to deal with graded syntactic acceptability [116] and sequence probability, they should be grounded in considerations of descriptive parsimony, in links to semantics and form-meaning interactions, and they should not only account for production and perception, but also consider learnability and computational complexity (e.g. memory requirements, and accounting for limited memory in biological systems). Finally, formal models should be required to make predictions for empirical structures based on which they may be distinguished on an empirical basis.

One main topic in theoretical linguistics, orthogonal to the finer classifications within the CH, concerns the distinction between 'weak generative capacity' and 'strong generative capacity' that are relevant for debates in music and animal research. There are divergent definitions of these terms. Briefly, they concern the difference whether we can focus on just classes of sets of sequences (i.e. regarding models just by the surface sequences they model: 'weak generative capacity'), or need to look at classes of structural analyses that different models can generate (i.e. regarding models in terms of the hidden underlying structure they assume to model sequences: 'strong generative capacity').

This is relevant, for instance, for distinguishing issue (2) long-distance dependencies, from (3) context-freeness,

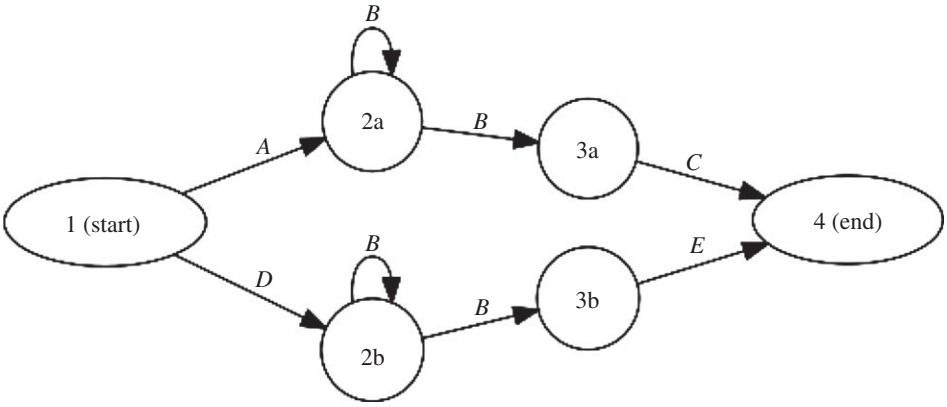

Figure 4. A finite-state automaton generating a repertoire consisting of two sequences: AB^nC and DB^nE (with $n > 0$). Note that the finite-state automaton is redundant in the way that it contains multiple instances of the same structure B^n.

above. A long-distance dependency in itself is not enough to prove the inadequacy of finite-state models (as we stated above); context-freeness is necessary only when long-distance dependencies can be embedded within (unboundedly many) other long-distance dependencies. For instance, when a bird sings songs of the structure AB^nC and DB^nE, we observe a long-distance dependency between A and C and between D and E, but the songs can be easily modelled with finite-state automata (figure 4) by just assuming two different (hidden) states from which the Bs are generated: one for the condition starting with A and ending with C, and one for the other. This explains why some efforts to empirically demonstrate context-freeness of bird song or music may not be unconvincing from a formal language theory perspective if they are based on just demonstrating a long-distance dependency. However, a long-distance dependency does have consequences for the underlying model that can be assumed for its strong generative capacity, representational capacity and compressive power: in the example shown in figure 4, we were forced to duplicate the state responsible for generating B, in fact we require $2m$ states (where m is the number of non-local dependency pairs, such as $A \ldots C$ or $D \ldots E$, that need to be encoded). Therefore, if there are multiple (finite), potentially nested non-local dependencies the number of required states grows exponentially (see also the comparable argument regarding the implicit acquisition of such structures in references [99,114]).

On similar grounds, one may argue that human musical capacities exceed not only Markovian, but probably also a finite-state representation (which is not a relevant or probable model here on the grounds just presented), based on empirical evidence that a recent study provided: human non-musicians were found to process non-local syntactic dependencies resulting from one level of centre-embedding in ecological music excerpts [117]. Comparable findings in animal vocalization are still missing.

This example illustrates that the CH as theoretical construct is irrelevant here in choosing the best model. If the intervening material in a long-distance dependency is very variable, even if not technically unbounded, considerations of parsimony, strong-generative capacity, elegant structure-driven compression and considerations of efficiency provide strong reasons to prefer a model other than the minimally required class in the CH or a different type of model altogether. Further, empirical testability and evaluation, for example in terms of Bayesian model comparison, come to play an important role in this context.

Another interesting situation where the distinction between weak and strong generative power matters arises when we combine a (finite-state/hidden Markov model (HMM, see below)) model of how elements are sequenced to create a song with a (finite-state/HMM) model of how songs are sequenced to create a song bout. As a simple example, consider a hypothetical bird that sings two different songs, A and B, each built up from some repertoire of elements, but showing a typical structure: song A might start with some slow introductory elements, and then continue with faster elements, whereas song B might start with fast, high-pitched elements, and continue with low-pitched ones. We can also imagine that the bird mainly uses song A in one context, and song B in another context. In such a case, the appropriate model might be a *hierarchical hidden Markov model*, which at the level of sequences-of-songs distinguishes different 'states' for different contexts, and at the level of sequences-of-elements distinguishes different states for different sections (start, end) of the song. Such a model is still finite-state in terms of weak generative power—because ultimately it can be described as involving a finite number (two contexts $\times$ two sections $=$ four states) of states, but it would be problematic to describe the model without using an account that represents the individual songs that are generated; accordingly, a model for describing song production in terms of strong generativity may potentially rather employ a formalism such as CFGs.

Other motivations to move beyond the confinements of the CH lie in the modelling of real-world structures that undermine some of the assumptions of the CH. Generally, the aspect that music involves not only multiple parallel streams of voices but also correlated streams of different features and complex timing constitutes a theme that received considerable research in the domain of music cognition, yet does not easily match with the principles that underlie the CH that is based on modelling a single sequence of words. In their cognitive model that is not only constrained to the case of music, Conklin & Witten [39] proposed successful extensions of n-gram models by combining n-gram models over different features and combined feature-spaces to optimize prediction based on the features that embody most information. They also combined predictions derived from the current piece (short-term model) with predictions derived from a corpus (long-term

model). Extending this framework, the IDyOM model [40] includes a search for progressively more information-theoretically efficient representations, which are shown in turn to give rise to progressively better predictors of human expectations. The IDyOM model has been shown to be successful in the domains of music and language [56,118]. Recent modelling approaches generalized the notion of modelling parallel feature streams into dynamic Bayesian networks that combine the advantages of hidden Markov models with modelling feature streams [42,119–121].

7. Dealing with noisy data: adding probabilities

The original CH is based on a number of assumptions that turn problematic in the light of ecological data. One main problem that is particularly relevant in the domain of music or animal songs is that the notion of grammaticality or well-formedness, which is fundamental for establishing and testing symbolic rules, is much less clear than in language (where there also are problems—see [122] for an insightful discussion). Most discussions of grammatical structure in music and animal song are based on so-called positive data (i.e. examples conforming with the proposed rules) and it is significantly more difficult to establish the validity and extension of rules in the absence of negative data (i.e. where humans or animals explicitly reject a malformed sequence). It is not clear whether ungrammatical/irregular structures in music are clear-cut or distinguished in agreement by non-expert or expert subjects. This difference between the linguistic and the musical case may also in part be explained by the fact that (at least in Western music), there is large divergence between active and passive musical interaction. Furthermore, negative data are, particularly in the case of animal research, more difficult to obtain and also less clear-cut or potentially graded rather than binary.

This issue motivates a number of changes in the nature of models. Models may be grounded in foundations other than grammaticality, such as optimal prediction or compression (e.g. predictive power [56], minimum description length [123], Bayesian model comparison), strong generativity or by motivation to express form of a higher-order structure (such as semantics or musical tension; see the following section). Another important way to build better models of cognition and deal with the issues above comes from employing syntactic gradience and reintroducing the probabilities that Chomsky abandoned along with his rejection of finite-state models. Apart from a large number of recent probabilistic models (such as the ones mentioned in the previous section) that go beyond the framework of the CH, it turns out that a hierarchy of probabilistic grammars can be defined that is analogous to the classical (and extended) CH and exhibits the same expressive power, with the additional advantage that grammars from this hierarchy can straightforwardly deal with noisy data and frequency effects and lend themselves to information theoretic methodologies such as model comparison, compression or minimum description length [124,125]. *n*-Gram models constitute the probabilistic counterpart of SLTLs, whereas the probabilistic counterpart of finite-state automata are hidden Markov models; CFGs can be straightforwardly extended to probabilistic CFGs.

Hidden Markov models constitute one type of model that has been very successful in all domains of language, music and animal songs. Comprehensively reviewed by Rabiner [126], the HMM assumes a number of underlying (hidden) states each of which emits surface symbols from given probabilistic emission vectors, a Markov matrix defining transition probabilities between states (including remaining in the same state) and a probability vector modelling the start state. HMMs have been very successful in modelling language, music and animal songs [42,49,59,127,128].

Thanks to the probabilities, we can talk about degrees of fit, and thus select models in a Bayesian model comparison paradigm that have the highest posterior probability given the degree of fit and prior beliefs; also, the probabilistic grammar framework does not require wellformedness as a criterion, but rather can use the likelihood of observing particular sentences, songs or musical structure as a criterion [8].

8. Dealing with meaning: adding semantics

The CH of formal grammars has its limitations, but has played a major role in generating hypotheses to test, not only on natural language, but also on animal songs and music. But where does this leave semantics? Chomsky's original work stressed the independence of syntax from semantics, but that does not mean that semantics is not important for claims about human uniqueness, even for linguists working within a 'Chomskian' paradigm. Berwick *et al.* [7], for instance, use the point that birdsong crucially lacks underlying semantic representations to argue against the usefulness of bird song as a comparable model system for human language. The reason why this is so is that in natural language, the transfinite-state structure is not some idiosyncratic feature of the word streams we produce, but something that plays a key role in mediating between thought (the conceptual–intentional system in Chomsky's terms) and sound (the phonetic articulatory–perceptual system).

Crucially, the conceptual–intentional system is also a hierarchical, combinatorial system (most often modelled using some variety of symbolic logic). From that perspective, grammars from the (extended) CH describe only one half of the system; a full description of natural language would involve a transducer that maps meanings to forms and vice versa [49,129]. For instance, finite-state grammars can be turned into finite-state transducers, and CFGs into synchronous CFGs. All the classes of grammars in the CH have a corresponding class of transducers (see Knight & Graehl [130] for an overview). Depending on the type of interaction we allow between syntax and semantics, there might or might not be consequences for the set of grammatical sentences that a grammar allows if we extend the grammar with semantics. But, the extension is, in any case, relevant for assessing the adequacy of the combined model—for example we can ask whether a particular grammar supports the required semantic analysis—as well as for determining the likelihood of sentences and alternative analyses of a sentence.

Do we need transducers to model structure building in animal songs and music? There have been debates about forms of musical meaning and its neurocognitive correlates. However, a large number of researchers in the field agree that music may feature simple forms of associative meaning and connotations as well as illocutionary forms of expression, but lacks kinds of more complex forms of combinatorial semantics (see the discussion of references [131–137]).

However, it is possible, as mentioned above, to conceive of complex forms of musical tension that involve nested patterns of expectancy and prolongation as an abstract secondary structure that motivates syntactic structures at least in Western tonal music and in analogy would require characterizing a transducer mapping syntactic structure and corresponding structures of musical tension in future research.

There have similarly been debates about the semantic content of animal communication. There are a few reported cases of potential compositional semantics in animal communication (cf. [138]), but these concern sequences of only two elements and thus do not come close to needing the expressiveness of finite-state or more complex transducers. For all animal vocalizations that have non-trivial structure, such as the songs of nightingales [139], blackbirds [8,108], pied butcherbirds [140] or humpback whales [141,142], it is commonly assumed that there is no combinatorial semantics underlying it. However, it is important to note that the ubiquitous claim that animal songs do not have combinatorial, semantic content is actually based on little to no experimental data. As long as the necessary experiments are not designed and performed, the absence of evidence of semantic content should not be taken as evidence of absence.

If animal songs do indeed lack semanticity, they would be more analogous to human music than to human language. The analogy to music would then not primarily be based on the surface similarity to music on the level of the communicative medium (use of pitch, timbre, rhythm or dynamics), but on functional considerations such that they do not constitute a medium to convey types of (propositional) semantics or simpler forms of meaning, but are instances of comparably free play with form and displays of creativity (see below and Wiggins *et al.* [143]).

Does this view on music–animal song analogies have any relevance for the study of language? There are reasons to argue it does, because music and human language may be regarded as constituting a continuum of forms of communication that is distinguished in terms of specificity of meaning [144–146])—consider, for instance, several forms of language that may be considered closer to a 'musical use' in terms of their use of pitch, rhythm, metre, semantics, for example motherese, prayers, mantras, poetry, nursery rhymes, forms of the utterance 'huh' (see [147]), etc. Drawing a strict dichotomy between music and language may further be a strongly anthropomorphic distinction that may have little match in animal communication. Animal vocalizations may be motivated by forms of meaning (that are not necessarily comparable with combinatorial semantics), for example, expressing aggression or submission, warning of predators, group cohesion, social contagion or may constitute free play of form for display of creativity, for instance (but not necessarily) in the context of reproduction. Given that structure and structure building moving from the language end to the music end is less constrained by semantic forms, more richness of structural play and creativity is expected to occur on the musical side [143].

9. Dealing with gradations: adding continuous-valued variables

A final move to a new class of successful models originates in a far research extension of the CH framework where the categorical symbols used in rewrite grammars are replaced by vectors. Thus, instead of having a rule $X \rightarrow YZ$, where X, Y and Z are categorical objects (such as a 'prepositional phrase' (PP) in linguistics, or a motif in zebra finch song), we treat X, Y and Z as n-dimensional vectors of numbers (which could be binary, integer, rational or real numbers; for example [0,1,0,0,1, . . .] or [0.45,0.3333,0.96, . . .]). By choosing vectors at maximum distance from each other such grammars can perfectly 'mimic' classical, symbolic grammars (which now become the ends of a continuum—and are still important as potential attractors in learning and evolution); but, additionally, vector grammars offer a natural way to model similarity between phrase types (assuming some noisy reading of the vectors, making correct recognition of a vector less likely the further it is from its prototype; see also reference [148], for an insightful discussion of the relation between simple recurrent networks and formal grammars).

A simple example (discussed in reference [149]) is the category of the word 'near', which sometimes acts like an adjective (ADJ, 'the future is near') and sometimes like a preposition (PREP, 'the house is near the river'). In standard symbolic as well as probabilistic grammars, this is normally modelled by having two entries in the lexicon, one with category ADJ, and the other with category PREP. Interestingly, however, 'near' shows behaviour inconsistent with this simple duplication approach: as a preposition, it inherits a property of adjectives, namely that it can be used as a comparative (as in 'the house is nearer the river than the road'). In vector grammars, such cases can be dealt with by assigning 'near' a vector in between the cluster of prepositions and the cluster of adjectives.

In computational linguistics, vector grammars (which have a close relation to neural networks models of linguistic structure from the 1990s [150,151]) are experiencing a new wave of excitement following some successes with learning such grammars from data for practical natural language processing tasks [152–155]. These successes are due not only to the ability to represent gradedness but also (and perhaps more so) to the fact that models with continuous values allow gradual learning procedures (in these papers, the procedure is 'backpropagation through structure' [156]).

While vector grammars have, to the best of our knowledge, not been applied yet to music and animal vocalizations, we expect that they offer much potential in these fields. Efforts to apply symbolic grammars in these domains often encounter difficulties with dealing with phenomena that seem fundamentally continuous in nature, such as loudness and pitch variation, beat, etc. Musical examples would exhibit micro-differences in pitch corresponding with syntactic function (e.g. raised leading notes), complex note features (such as slides used in traditional North Indian rāgās) or timbre features in distorted guitar solos that may correspond with syntactic function, such as stress, intonation or timbre cues may be modified to direct attention towards elements with important syntactic functions (e.g. stability or marking departures at constituent or phrase boundaries; further empirical research is, however, required to examine such connections). Altogether vector grammars may constitute a well-suited formalism to capture the use of noisy, non-tonal and complex sounds in music across cultures that may be challenging for traditional symbolic models.

One interesting feature of vector grammars is that they are computationally more expressive than their symbolic counterparts—in fact, they result in an entirely different

class of complexity that is not comparable with the models of the CH from which the models originated (this is due to the large amounts of information that the large real-valued vectors may carry). For instance, Rodriguez [157] demonstrated that a right-branching vector grammar (or, in fact, its special case, the Elman network [150]) can generate several context-free languages, even though it is well-known that right-branching symbolic grammars are finite-state. The reason is, Rodriquez argues, that small displacements in a continuous vector space can be used to implement a counter. We may expect that also other classes of symbolic grammars can be approximated by vector grammars with very simple structure (even if the non-terminals are in some sense more complex). While much theoretical work exploring their expressive power is still necessary, they provide another motivation to move on to probabilistic, non-symbolic models that move beyond the constraints of the CH.

In short, vector grammars provide a generalization over the classical symbolic grammars from the extended CH. In particular, Rodriquez's [157] results demonstrate a continuum between finite-state and (at least some) context-free languages might be used to call into question the *a priori* plausibility of the hypothesis that context-freeness is uniquely human (contra Fitch & Hauser [158])—a hypothesis only originating from the reification of the CH.

Following this line of reasoning, one could even argue that previous accounts have been overstated in their focus on 'architectural' constraints on structure building operations when discussing whether 'animals may not be able to process context-free languages, whereas humans can' or whether 'music has or has no access to the recursion operation' (see also the discussion in reference [159]). One may hypothesize, then, that the basic toolbox of structure building operations between humans, primates and birds is largely shared, but that music, language, primate and bird calls and bird and cetacean songs have recruited (perhaps in a process of cultural evolution) very different operations from that toolbox because they serve very different functions—a thought that we explore in §10.

10. Function, motivation and context

Our journey through the computational models of structure building—from Shannon's *n*-grams, via the CH to vector grammars and models beyond the CH—has thus uncovered many useful models for how sequences of sound might be generated and processed, which also motivate that a richer space of possibility may need to be explored before potential explanations may resort to the assumption of hard biological constraints. What could be alternative explanations for some of the differences and similarities in structure that we do observe in these domains? In this final section, we consider explanations based on function, motivation and context.

Fundamental to a cross-cultural concept of music is the form of joint group interaction: music may be a form to coordinate and synchronize a large group in terms of joint action, production and perception [145]. Music may be monological, dialogical, responsorial (individual–group) or an entire joint group-based activity [19]. This view makes it possible to understand the principles of structure building, its flexibility and constraints based on affording these forms of interactive communication. In this context, the notion of 'absolute music' [160] as a highly complex form of Western art is a very small subset of music of the world and a construct that constitutes an exception in an overarching cross-cultural comparison (cf. [145,161]). As mentioned above, forms of high musical complexity may stem in part from traditions of formal scholarly teaching and the use of sophisticated notational systems and probably can be correlated to uses of music in a 'monological' form rather than a 'joint group-based activity' (even when an orchestra is performing *monologically* for a larger audience). Further, theories of music that take such examples of absolute Western art music as basis for *general* claims of musical complexity and processing and specifically linking findings about complexity and recursion [37] to the debate concerning the human language faculty [28,162,163] need to be examined carefully in order to establish the extent to which the claims generalize to music across cultures (see reference [145] and the discussion in this review), because this link is not given *per se*. Nonetheless, observations concerning complexity in Western tonality and corresponding processing requirements may contribute important evidence towards the complexity that music cognition may reach (in terms of proofs of existence). While the considerations above do not invalidate comparisons of structural/syntactic complexity in animal vocalization, human language and music, it may further be the case that the conditions of communication and the social context of music making (such as joint group activity) may impose limits on the structural complexity that may be used. Therefore, these considerations ask for careful examination of the underlying conditions of communication.

Once the purpose of communication is not necessarily required to be grounded in propositional semantics, the question regarding the foundation of the form of communication becomes an interesting theoretical challenge (cf. [164]). From a cross-cultural perspective, music may be regarded as a tool for dealing with situations of social function and social uncertainty [165], it may be construed as a form of display, communication and regulation of emotion and tension [19,166] or an (abstract) play with structure *per se* [167,168]. These different forms to construct music have different implications on structure (and, of course, they are not mutually exclusive): music in social and interactive contexts is required to afford rhythmic entrainment and to build on socially established conventions of structures. The notion of music as carrier or inducer of emotions implies stronger constraints on musical structure, namely those the processing of which results in triggering the corresponding emotional mixture. In contrast, construing music as (mere) free play of form and formal complexity entails the least formal constraints.

In the above-mentioned context, many conditions of animal vocalization appear to match features of human music; as far as we can tell they are not governed by complex semantics, they are interactive and group activities but can also occur alone. Accordingly, principles of structure building in music and animal vocalization may be considerably different from the ones governing language and be constrained by different communicative motivations than the ones found in language. While simple forms of expressions such as fear or aggression encoded in prosodic structure and illocutionary speech acts may be similar in all three domains, this functional analysis suggests that syntactic structures may not necessarily converge in these domains as mere structural comparison as proposed by Jackendoff & Lerdahl [28] (see also [163]).

11. Conclusion

In this review, we have discussed different formal models of syntactic structure building, building blocks and functional motivations of structure in language, music and birdsong. With these considerations, we aim to lay a common ground for future fruitful formal analysis, shared methodologies and comparative formal and empirical research addressing common cognitive questions that were raised in these domains. Not many steps have been made to bring theoretical approaches in music and animal vocalization into common terms that can be compared with approaches established in formal and computational linguistics. These questions are fundamental to understand commonalities and differences between humans and animal cognition as well as between music and language, responding not only to Hauser, Chomsky and Fitch's provocative hypothesis concerning the 'exceptional' role of the human cognitive/communicative abilities [11].

Music and language may constitute a continuum in the human communicative toolkit [169,170] and forms of animal vocalizations may relate to different aspects in this spectrum (and not necessarily be related to only one of them). The level of analogy may not necessarily be based on formal considerations, but may also depend on corresponding semantics (or second tier structure), context, function and the underlying motivation of communication.

We discussed several classes of formal models that lend themselves to computational implementation and empirical evaluation. While the (extended) CH provides a useful perspective for the comparison of different theoretical approaches and predictions concerning properties of structures, it is important to bear in mind the limitations we discussed and the fact that cognitively plausible mechanisms may not be well represented in terms of the CH. The notion of grammatical gradience and other factors such as the necessary flexibility of dealing with noise and uncertainty in ecological structures motivate extensions towards probabilistic models. Finally, we have discussed recent models that add gradation and continuous-valued variable spaces. Such models link formal grammar and neural network approaches and add the power to deal with ecological structures that are inherently continuous. Finally, they predict that a single cognitively relevant framework may predict sequences of different complexity in the CH and therefore have the potential to undermine assumptions concerning categorically different cognitive capacities between human and animal forms of communication [11,171].

Acknowledgements. We thank four anonymous reviewers for very useful advice and detailed comments in assembling this review. This review was made possible under the support of the Lorentz Centre, Leiden, through their workshop on *Cognition, biology and the origins of musicality*.

Funding statement. M.R. was generously supported by the MIT Department of Linguistics and Philosophy as well as the Zukunftskonzept at TU Dresden supported by the Exzellenzinitiative of the Deutsche Forschungsgemeinschaft. W.Z. is supported by the New Generation Initiative of the Faculty of Humanities at the University of Amsterdam. G.A.W. is supported by the Lrn2Cre8 and ConCreTe projects, which acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant numbers 610859 and 611733. C.S. acknowledges funding from the Excellence Clusters Languages of Emotions and Neurocure, the collaborative research grant SFB 665, and Bernstein focus area 'neural basis of learning', project 'variable tunes: neural mechanisms underlying learning motor sequences (01GQ0961/TP1, BMBF).

Endnotes

[1]An analogue to the notion of repertoire size in bird song may also potentially turn fruitful for music research.

[2]No research studies have so far addressed the question whether the power of CFGs to express counting and numbered relationships between elements in musical or animal song sequences are required in ecological real-world materials.

[3]We thank reviewer 1 for her/his insightful points on the shortcomings of existing approaches applying formal language theory to empirical data from linguistic, music and animal vocalization research and on requirements for better models.

References

1. Doupe AJ, Kuhl PK. 1999 Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631. (doi:10.1146/annurev.neuro.22.1.567)
2. Bolhuis JJ, Okanoya K, Scharff C. 2010 Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* **11**, 747–759. (doi:10.1038/nrn2931)
3. Hurford JR. 2011 *The origins of grammar: language in the light of evolution II*, vol. 2. Oxford, UK: Oxford University Press.
4. Knörnschild M, Feifel M, Kalko EK. 2014 Male courtship displays and vocal communication in the polygynous bat *Carollia perspicillata*. *Behaviour* **151**, 781–798. (doi:10.1163/1568539X-00003171)
5. Yip MJ. 2006 The search for phonology in other species. *Trends Cogn. Sci.* **10**, 442–446. (doi:10.1016/j.tics.2006.08.001)
6. Spierings MJ, ten Cate C. 2014 Zebra finches are sensitive to prosodic features of human speech. *Proc. R. Soc. B* **281**, 20140480. (doi:10.1098/rspb.2014.0480)
7. Berwick RC, Okanoya K, Beckers GJ, Bolhuis JJ. 2011 Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* **15**, 113–121. (doi:10.1016/j.tics.2011.01.002)
8. ten Cate C, Lachlan R, Zuidema W. 2013 Analyzing the structure of bird vocalizations and language: finding common ground. In *Birdsong, speech, and language: exploring the evolution of mind and brain* (eds JJ Bolhuis, B Everaert), pp. 243–260. Cambridge, MA: MIT Press.
9. Markowitz JE, Ivie E, Kligler L, Gardner TJ. 2013 Long-range order in canary song. *PLoS Comput. Biol.* **9**, e1003052. (doi:10.1371/journal.pcbi.1003052)
10. Sasahara K, Cody ML, Cohen D, Taylor CE. 2012 Structural design principles of complex bird songs: a network-based approach. *PLoS ONE* **7**, e44436. (doi:10.1371/journal.pone.0044436)
11. Hauser MD, Chomsky N, Fitch WT. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)
12. Fitch W. 2005 The evolution of music in comparative perspective. *Ann. N.Y. Acad. Sci.* **1060**, 29–49. (doi:10.1196/annals.1360.004)
13. Fitch W. 2006 The biology and evolution of music: a comparative perspective. *Cognition* **100**, 173–215. (doi:10.1016/j.cognition.2005.11.009)
14. Rothenberg D, Roeske TC, Voss HU, Naguib M, Tchernichovski O. 2014 Investigation of musicality in birdsong. *Hear. Res.* **308**, 71–83. (doi:10.1016/j.heares.2013.08.016)
15. Hockett CF. 1960 The origin of speech. *Sci. Am.* **203**, 88–111. (doi:10.1038/scientificamerican0960-88)
16. Adger D. 2003 *Core syntax: a minimalist approach*, vol. 33. Oxford, UK: Oxford University Press.
17. Bloom P. 2004 Can a dog learn a word. *Science* **304**, 1605–1606. (doi:10.1126/science.1099899)
18. Scharff C, Petri J. 2011 Evo-devo, deep homology and FoxP2: implications for the evolution of

speech and language. *Phil. Trans. R. Soc. B* **366**, 2124–2140. (doi:10.1098/rstb.2011.0001)

19. Brown S, Jordania J. 2013 Universals in the world's musics. *Psychol. Music* **41**, 229–248. (doi:10.1177/0305735611425896)
20. Trehub SE. 2000 Human processing predispositions and musical universals. In *The origins of music* (eds N Wallin, B Merker, S Brown), pp. 427–448. Cambridge, MA: MIT Press.
21. Fabb N, Halle M. 2012 Grouping in the stressing of words, in metrical verse, and in music. In *Language and music as cognitive systems* (eds P Rebuschat, M Rohmeier, J Hawkins, I Cross), pp. 4–21. Oxford: Oxford University Press.
22. Narmour E. 1990 *The analysis and cognition of basic melodic structures: the implication realization model*. Chicago, IL: University of Chicago Press.
23. Narmour E. 1992 *The analysis and cognition of melodic complexity: the implication-realization model*. Chicago, IL: University of Chicago Press.
24. Cross I. 2012 Cognitive science and the cultural nature of music. *Top. Cogn. Sci.* **4**, 668–677. (doi:10.1111/j.1756-8765.2012.01216.x)
25. Margulis EH. 2014 *On repeat: how music plays the mind*. Oxford, UK: Oxford University Press.
26. Huron DB. 2006 *Sweet anticipation: music and the psychology of expectation*. Cambridge, MA: MIT Press.
27. Widdess DR. 1981 Aspects of form in North Indian ālāp and dhrupad. In *Music and tradition: essays on Asian and other musics presented to Laurence Picken* (eds DR Widdes, RF Wolpert), pp. 143–182. Cambridge, UK: Cambridge University Press.
28. Jackendoff R, Lerdahl F. 2006 The capacity for music: what is it, and what's special about it? *Cognition* **100**, 33–72. (doi:10.1016/j.cognition.2005.11.005)
29. Rohrmeier MA, Koelsch S. 2012 Predictive information processing in music cognition. A critical review. *Int. J. Psychophysiol.* **83**, 164–175. (doi:10.1016/j.ijpsycho.2011.12.010)
30. Tymoczko D. 2006 The geometry of musical chords. *Science* **313**, 72–74. (doi:10.1126/science.1126287)
31. Callender C, Quinn I, Tymoczko D. 2008 Generalized voice-leading spaces. *Science* **320**, 346–348. (doi:10.1126/science.1153021)
32. Quinn I, Mavromatis P. 2011 Voice-leading prototypes and harmonic function in two chorale corpora. In *Mathematics and computation in music* (eds C Agon, M Andreatta, G Assayag, E Amiot, J Bresson, J Mandereau), pp. 230–240. Berlin, Germany: Springer.
33. Aldwell E, Schachter C, Cadwallader A. 2010 *Harmony and voice leading*. Boston, MA: Cengage Learning.
34. Winograd T. 1968 Linguistics and the computer analysis of tonal harmony. *J. Music Theory* **12**, 2–49. (doi:10.2307/842885)
35. Rohrmeier M. 2007 A generative grammar approach to diatonic harmonic structure. In *Proc. Fourth Sound and Music Computing Conf.* (eds H Spyridis, A Georgaki, C Anagnostopoulou, G Kouroupetroglou), pp. 97–100. Athens, Greece: National and Kapodistrian University of Athens.
36. Rohrmeier M. 2011 Towards a generative syntax of tonal harmony. *J. Math. Music* **5**, 35–53. (doi:10.1080/17459737.2011.573676)
37. Lerdahl F, Jackendoff R. 1983 *A generative theory of tonal music*. Cambridge, MA: MIT Press.
38. Rohrmeier M, Neuwirth M. 2014 Towards a syntax of the classical cadence. In *What is a cadence? theoretical and analytical perspectives on cadences in the classical pepertoire* (eds M Neuwirth, P Bergé), pp. 285–336. Leuven, Belgium: Leuven University Press.
39. Conklin D, Witten IH. 1995 Multiple viewpoint systems for music prediction. *J. New Music Res.* **24**, 51–73. (doi:10.1080/09298219508570672)
40. Pearce M. 2005 *The construction and evaluation of statistical models of melodic structure in music perception and composition*. London, UK: City University.
41. Whorley RP, Wiggins GA, Rhodes C, Pearce MT. 2013 Multiple viewpoint systems: time complexity and the construction of domains for complex musical viewpoints in the harmonization problem. *J. New Music Res.* **42**, 237–266. (doi:10.1080/09298215.2013.831457)
42. Rohrmeier M, Graepel T. 2012 Comparing feature-based models of harmony. In *Proc. Ninth Int. Symp. on Computer Music Modelling and Retrieval* (eds R Kronland-Martinet, S Ystad, M Aramaki, M Barthet, S Dixon), pp. 357–370.
43. Odom K, Hall M, Riebel K, Omland K, Langmore N. 2014 Female song is widespread and ancestral in songbirds. *Nat. Commun.* **5**, 3379. (doi:10.1038/ncomms4379)
44. Marler P. 2004 *Bird calls: a cornucopia for communication*, pp. 132–177. San Diego, CA: Elsevier.
45. Honing H, Ploeger A. 2012 Cognition and the evolution of music: pitfalls and prospects. *Top. Cogn. Sci.* **4**, 513–524. (doi:10.1111/j.1756-8765.2012.01210.x)
46. Tierney AT, Russo FA, Patel AD. 2011 The motor origins of human and avian song structure. *Proc. Natl Acad. Sci. USA* **108**, 15 510–15 515. (doi:10.1073/pnas.1103882108)
47. Amador A, Margoliash D. 2013 A mechanism for frequency modulation in songbirds shared with humans. *J. Neurosci.* **33**, 11 136–11 144. (doi:10.1523/JNEUROSCI.5906-12.2013)
48. Shannon CE. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 and 623–656. (doi:10.1002/j.1538-7305.1948.tb01338.x)
49. Jurafsky D, Martin JH. 2000 *Speech & language processing*. New Delhi, India: Pearson Education India.
50. Isaac D, Marler P. 1963 Ordering of sequences of singing behaviour of mistle thrushes in relationship to timing. *Anim. Behav.* **11**, 179–188. (doi:10.1016/0003-3472(63)90027-7)
51. Chatfield C, Lemon RE. 1970 Analysing sequences of behavioural events. *J. Theor. Biol.* **29**, 427–445. (doi:10.1016/0022-5193(70)90107-4)
52. Slater P. 1983 Bird song learning: theme and variations. *Perspect. Ornithol.* **12**, 475–499. (doi:10.1017/CBO9780511759994.014)
53. Okanoya K. 2004 The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. *Ann. N.Y. Acad. Sci.* **1016**, 724–735. (doi:10.1196/annals.1298.026)
54. Briefer E, Osiejuk TS, Rybak F, Aubin T. 2010 Are bird song complexity and song sharing shaped by habitat structure? An information theory and statistical approach. *J. Theor. Biol.* **262**, 151–164. (doi:10.1016/j.jtbi.2009.09.020)
55. Ames C. 1989 The Markov process as a compositional model: a survey and tutorial. *Leonardo* **22**, 175–187. (doi:10.2307/1575226)
56. Pearce MT, Wiggins GA. 2012 Auditory expectation: the information dynamics of music perception and cognition. *Top. Cogn. Sci.* **4**, 625–652. (doi:10.1111/j.1756-8765.2012.01214.x)
57. Lipkind D *et al*. 2013 Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* **498**, 104–108. (doi:10.1038/nature12173)
58. ten Cate C, Okanoya K. 2012 Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Phil. Trans. R. Soc. B* **367**, 1984–1994. (doi:10.1098/rstb.2012.0055)
59. 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)
60. Katahira K, Suzuki K, Kagawa H, Okanoya K. 2013 A simple explanation for the evolution of complex song syntax in Bengalese finches. *Biol. Lett.* **9**, 20130842. (doi:10.1098/rsbl.2013.0842)
61. Krumhansl CL. 2004 The cognition of tonality: as we know it today. *J. New Music Res.* **33**, 253–268. (doi:10.1080/0929821042000317831)
62. Krumhansl CL, Kessler EJ. 1982 Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychol. Rev.* **89**, 334–368. (doi:10.1037/0033-295X.89.4.334)
63. Schellenberg EG. 1997 Simplifying the implication-realization model of melodic expectancy. *Music Percept.* **14**, 295–318. (doi:10.2307/40285723)
64. Schellenberg EG. 1996 Expectancy in melody: tests of the implication-realization model. *Cognition* **58**, 75–125. (doi:10.1016/0010-0277(95)00665-6)
65. Krumhansl CL. 1995 Music psychology and music theory: problems and prospects. *Music Theory Spectr.* **17**, 53–80. (doi:10.2307/745764)
66. Eerola T. 2003 *The dynamics of musical expectancy. Cross-cultural and statistical approaches to melodic expectations*. University of Jyväskylä, Finland.
67. Piston W. 1948 *Harmony*. New York, NY: W.W.Norton & Company.
68. Rameau JP. 1971 *Treatise on harmony*. Translated by Philip Gossett. New York, NY: Dover.
69. Hedges T, Rohrmeier M. 2011 Exploring Rameau and beyond: a corpus study of root progression theories. In *Mathematics and computation in music. Lecture notes in artificial intelligence* (*6726*) (eds C Agon, E Amiot, M Andreatta, G Assayag, J Bresson, J Mandereau), pp. 334–337. Berlin, Germany: Springer.

70. Temperley D. 2001 *The cognition of basic musical structures*. Cambridge, MA: MIT Press.
71. Pearce MT, Wiggins GA. 2006 Expectation in melody: the influence of context and learning. *Music Percept.* **23**, 377–405. (doi:10.1525/mp.2006.23.5.377)
72. Rohrmeier M. 2014 Musical expectancy. Bridging music theory, cognitive and computational approaches. *Zeitschrift der Gesellschaft für Musiktheorie.* **10**. See http://www.gmth.de/zeitschrift/artikel/724.aspx.
73. Ponsford D, Wiggins G, Mellish C. 1999 Statistical learning of harmonic movement. *J. New Music Res.* **28**, 150–177. (doi:10.1076/jnmr.28.2.150.3115)
74. Reis BY. 1999 *Simulating music learning with autonomous listening agents: entropy, ambiguity and context*. Cambridge, UK: Computer Laboratory, University of Cambridge.
75. Rohrmeier M. 2005 *Towards modelling movement in music: Analysing properties and dynamic aspects of pc set sequences in Bachs chorales*. Master's thesis, University of Cambridge, UK.
76. Pearce MT *et al*. 2010 The role of expectation and probabilistic learning in auditory boundary perception: a model comparison. *Perception* **39**, 1365–1391. (doi:10.1068/p6507)
77. Pearce M, Wiggins G. 2004 Improved methods for statistical modelling of monophonic music. *J. New Music Res.* **33**, 367–385. (doi:10.1080/0929821052000343840)
78. Chomsky N. 1956 Three models for the description of language. *IRE Trans. Inf. Theory* **2**, 113–124. (doi:10.1109/TIT.1956.1056813)
79. Keiler A. 1978 Bernstein's 'the unanswered question' and the problem of musical competence. *Music. Q.* **64**, 195–222. (doi:10.1093/mq/LXIV.2.195)
80. Keiler A. 1983 On some properties of Schenker's pitch derivations. *Music Percept.* **1**, 200–228. (doi:10.2307/40285256)
81. Steedman MJ. 1984 A generative grammar for jazz chord sequences. *Music Percept.* **2**, 52–77. (doi:10.2307/40285282)
82. Steedman M. 1996 The blues and the abstract truth: music and mental models. In *Mental models in cognitive science* (eds A Garnham, J Oakhill), pp. 305–318. Mahwah, NJ: Erlbaum.
83. de Haas B, Rohrmeier M, Veltkamp RC, Wiering F. 2009 Modeling harmonic similarity using a generative grammar of tonal harmony. In *Proc.10th Intl. Soc. Music Information Retrieval Conf.* (*ISMIR 2009*), pp. 549–554.
84. Granroth-Wilding M, Steedman M. 2014 A robust parser-interpreter for jazz chord sequences. *J. New Music Res*. **43**, 355–374. (doi:10.1080/09298215.2014.910532).
85. Schenker H. 1935 *Der Freie Satz. Neue musikalische Theorien und Phantasien*. Liège, Belgium: Margada.
86. Marsden A. 2010 Schenkerian analysis by computer: a proof of concept. *J. New Music Res.* **39**, 269–289. (doi:10.1080/09298215.2010.503898)
87. Temperley D. 2011 Composition, perception, and Schenkerian theory. *Music Theory Spectr.* **33**, 146–168. (doi:10.1525/mts.2011.33.2.146)
88. Hofstadter DH. 1980 *Gödel, Escher, Bach: an eternal golden braid: a metaphoric fugue on minds and machines in the spirit of Lewis Carroll*. London, UK: Penguin Books.
89. Lerdahl F, Krumhansl CL. 2007 Modeling tonal tension. *Music Percept*. **24**, 329–366. (doi:10.1525/mp.2007.24.4.329)
90. Lehne M, Rohrmeier M, Koelsch S. 2014 Tension-related activity in the orbitofrontal cortex and amygdala: an fMRI study with music. *Soc. Cogn. Affect. Neurosci.* **9**, 1515–1523. (doi:10.1093/scan/nst141)
91. Zengel MS. 1962 Literacy as a factor in language change. *Am. Anthropol.* **64**, 132–139. (doi:10.1525/aa.1962.64.1.02a00120)
92. Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. 2006 Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204–1207. (doi:10.1038/nature04675)
93. Abe K, Watanabe D. 2011 Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* **14**, 1067–1074. (doi:10.1038/nn.2869)
94. Van Heijningen CA, De Visser J, Zuidema W, Ten Cate C. 2009 Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl Acad. Sci. USA* **106**, 20 538–20 543. (doi:10.1073/pnas.0908113106)
95. Zuidema W. 2013 Context-freeness revisited. In *Proc. 35th Annu. Conf. Cognitive Science Society* (eds M Knauff, M Pauen, N Sebanz, I Wachsmuth), pp. 1664–1669. Austin, TX: Cognitive Science Society.
96. Beckers GJ, Bolhuis JJ, Okanoya K, Berwick RC. 2012 Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* **23**, 139–145. (doi:10.1097/WNR.0b013e32834f1765)
97. Jiang S, Zhu L, Guo X, Ma W, Yang Z, Dienes Z. 2012 Unconscious structural knowledge of tonal symmetry: tang poetry redefines limits of implicit learning. *Conscious. Cogn.* **21**, 476–486. (doi:10.1016/j.concog.2011.12.009)
98. Uddén J, Ingvar M, Hagoort P, Petersson KM. 2012 Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: investigating the push-down stack model. *Cogn. Sci.* **36**, 1078–1101. (doi:10.1111/j.1551-6709.2012.01235.x)
99. Rohrmeier M, Fu Q, Dienes Z. 2012 Implicit learning of recursive context-free grammars. *PLoS ONE* **7**, e45885. (doi:10.1371/journal.pone.0045885)
100. Rohrmeier M, Cross I. 2009 Tacit tonality: implicit learning of context-free harmonic structure. In *Proc. Seventh Triennial Conf. of European Society for the Cognitive Sciences of Music*. See http://urn.fi/URN:NBN:fi:jyu-2009411312.
101. Li F, Jiang S, Guo X, Yang Z, Dienes Z. 2013 The nature of the memory buffer in implicit learning: learning Chinese tonal symmetries. *Conscious. Cogn.* **22**, 920–930. (doi:10.1016/j.concog.2013.06.004)
102. Rohrmeier M, Rebuschat P. 2012 Implicit learning and acquisition of music. *Top. Cogn. Sci.* **4**, 525–553. (doi:10.1111/j.1756-8765.2012.01223.x)
103. Kuhn G, Dienes Z. 2005 Implicit learning of nonlocal musical rules: implicitly learning more than chunks. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 1417–1432. (doi:10.1037/0278-7393.31.6.1417)
104. Jäger G, Rogers J. 2012 Formal language theory: refining the Chomsky hierarchy. *Phil. Trans. R. Soc. B* **367**, 1956–1970. (doi:10.1098/rstb.2012.0077)
105. Tymoczko D, Meeùs N. 2003 Progressions fondamentales, fonctions, degrés: une grammaire de l'harmonie tonale élémentaire. *Musurgia* **10**, 35–64.
106. De Clercq T, Temperley D. 2011 A corpus analysis of rock 920 harmony. *Popular Music* **30**, 47–70. (doi:10.1017/S026114301000067X)
107. Rohrmeier M, Cross I. 2008 Statistical properties of tonal harmony in Bachs chorales. In *Proc. 10th Int. Conf. on music perception and cognition*, pp. 619–627. Citeseer.
108. Todt D. 1975 Social learning of vocal patterns and modes of their application in grey parrots (*Psittacus erithacus*) 1, 2, 3. *Zeitschrift für Tierpsychologie* **39**, 178–188. (doi:10.1111/j.1439-0310.1975.tb00907.x)
109. Kershenbaum A, Bowles AE, Freeberg TM, Jin DZ, Lameira AR, Bohn K. 2014 Animal vocal sequences: not the Markov chains we thought they were. *Proc. R. Soc. B* **281**, 20141370. (doi:10.1098/rspb.2014.1370)
110. Joshi AK. 1985 How much context-sensitivity is required to provide reasonable structural descriptions: tree-adjoining grammars. In *Natural language parsing: psycholinguistic, computational and theoretical perspectives* (eds D Dowty, L Karttunen, A Zwicky), pp. 206–350. New York, NY: Cambridge University Press.
111. Steedman M. 2000 *The syntactic process*. Cambridge, MA: MIT Press/Bradford Books.
112. Rohrmeier M, Dienes Z, Guo X, Fu Q. 2014 Implicit learning and recursion. In *Language and recursion* (eds F Lowenthal, L Lefebvre), pp. 67–85. Berlin, Germany: Springer.
113. Shieber SM. 1987 *Evidence against the context-freeness of natural language*. Germany, Berlin: Springer.
114. Rohrmeier M, Dienes Z, Guo X, Fu Q. 2014 Implicit learning and recursion. In *Language and recursion*, pp. 67–85. Springer.
115. Fitch WT, Friederici AD. 2012 Artificial grammar learning meets formal language theory: an overview. *Phil. Trans. R. Soc. B* **367**, 1933–1955. (doi:10.1098/rstb.2012.0103)
116. Aarts B. 2004 Modelling linguistic gradience. *Stud. Lang.* **28**, 1–49. (doi:10.1075/sl.28.1.02aar)
117. Koelsch S, Rohrmeier M, Torrecuso R, Jentschke S. 2013 Processing of hierarchical syntactic structure in music. *Proc. Natl Acad. Sci. USA* **110**, 15 443–15 448. (doi:10.1073/pnas.1300272110)
118. Wiggins GA. 2012 'I let the music speak': cross-domain application of a cognitive model of musical learning. In *Statistical learning and language acquisition* (eds P Rebuschat, J Williams), pp. 463–494. Amsterdam, The Netherlands: De Gruyter.
119. Murphy KP. 2002 *Dynamic Bayesian networks: representation, inference and learning*. Berkeley, CA: University of California.

120. Raczynski SA, Vincent E, Sagayama S. 2013 Dynamic Bayesian networks for symbolic polyphonic pitch modeling. *IEEE Trans. Audio, Speech Lang. Process.* **21**, 1830–1840. (doi:10.1109/TASL.2013.2258012)
121. Paiement JF. 2008 *Probabilistic models for music.* Ecole Polytechnique Fédérale Lausanne, Switzerland.
122. Abney S. 1996 Statistical methods and linguistics. In *The balancing act: combining symbolic and statistical approaches to language* (eds J Klavans, P Resnik), pp. 1–26. Cambridge, MA: MIT Press.
123. Mavromatis P. 2009 Minimum description length modelling of musical structure. *J. Math. Music* **3**, 117–136. (doi:10.1080/17459730903313122)
124. Grünwald PD. 2007 *The minimum description length principle*. Cambridge, MA: MIT Press.
125. MacKay DJ. 2003 *Information theory, inference, and learning algorithms*, vol. 7. Citeseer.
126. Rabiner L. 1989 A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**, 257–286. (doi:10.1109/5.18626)
127. Raphael C, Stoddard J. 2004 Functional harmonic analysis using probabilistic models. *Comput. Music J.* **28**, 45–52. (doi:10.1162/0148926041790676)
128. Mavromatis P. 2005 A hidden Markov model of melody production in Greek church chant. *Comput. Musicol.* **14**, 93–112.
129. Zuidema W. 2013 Language in nature: on the evolutionary roots of a cultural phenomenon. In *The language phenomenon*, pp. 163–189. Berlin, Germany: Springer.
130. Knight K, Graehl J. 2005 An overview of probabilistic tree transducers for natural language processing. In *Computational linguistics and intelligent text processing*, pp. 1–24. Berlin, Germany: Springer.
131. Koelsch S. 2011 Towards a neural basis of processing musical semantics. *Phys. Life Rev.* **8**, 89–105. (doi:10.1016/j.plrev.2011.04.004)
132. Slevc LR, Patel AD. 2011 Meaning in music and language: three key differences: comment on towards a neural basis of processing musical semantics by Stefan Koelsch. *Phys. Life Rev.* **8**, 110–111. (doi:10.1016/j.plrev.2011.05.003)
133. Fitch W, Gingras B. 2011 Multiple varieties of musical meaning: comment on towards a neural basis of processing musical semantics by Stefan Koelsch. *Phys. Life Rev.* **8**, 108–109. (doi:10.1016/j.plrev.2011.05.004)
134. Cross I. 2011 The meanings of musical meanings: comment on towards a neural basis of processing musical semantics by Stefan Koelsch. *Phys. Life Rev.* **8**, 116–119. (doi:10.1016/j.plrev.2011.05.009)
135. Davies S. 2011 Questioning the distinction between intra-and extra-musical meaning: comment on towards a neural basis for processing musical semantics by Stefan Koelsch. *Phys. Life Rev.* **8**, 114–115.
136. Reich U. 2011 The meanings of semantics: comment on towards a neural basis of processing musical semantics by Stefan Koelsch. *Phys. Life Rev.* **8**, 120–121. (doi:10.1016/j.plrev.2011.05.012)
137. Koelsch S. 2011 Transitional zones of meaning and semantics in music and language: reply to comments on Towards a neural basis of processing musical semantics. *Phys. Life Rev.* **8**, 125–128. (doi:10.1016/j.plrev.2011.05.011)
138. Arnold K, Zuberbühler K. 2012 Call combinations in monkeys: compositional or idiomatic expressions? *Brain Lang.* **120**, 303–309. (doi:10.1016/j.bandl.2011.10.001)
139. Weiss M, Hultsch H, Adam I, Scharff C, Kipper S. 2014 The use of network analysis to study complex animal communication systems: a study on nightingale song. *Proc. R. Soc. B* **281**, 20140460. (doi:10.1098/rspb.2014.0460)
140. Taylor H, Lestel D. 2011 The Australian pied butcherbird and the natureculture continuum. *J. Interdiscip. Music Stud.* **5**, 57–83.
141. Payne RS, McVay S. 1971 Songs of humpback whales. *Science* **173**, 585–597. (doi:10.1126/science.173.3997.585)
142. Payne K, Payne R. 1985 Large scale changes over 19 years in songs of humpback whales in Bermuda. *Zeitschrift für Tierpsychologie* **68**, 89–114. (doi:10.1111/j.1439-0310.1985.tb00118.x)
143. Wiggins GA, Tyack P, Scharff C, Rohrmeier M. 2015 The evolutionary roots of creativity: mechanisms and motivations. *Phil. Trans. R. Soc. B* **370**, 20140099. (doi:10.1098/rstb.2014.0099)
144. Brown S. 2000 The 'musilanguage' model of music evolution. In *The origins of music* (eds N Wallin, B Merke, S Brown), pp. 271–300. Cambridge, MA: MIT Press.
145. Cross I. 2012 Music and biocultural evolution. In *The cultural study of music: a critical introduction*, 2nd edn (ed. RM Clayton, T Herbert). London, UK: Routledge.
146. Cross I, Woodruff GE. 2009 Music as a communicative medium. *Prehist. Lang.* **11**, 77. (doi:10.1093/acprof:oso/9780199545872.003.0005)
147. Dingemanse M, Torreira F, Enfield N. 2013 Is huh? A universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS ONE* **8**, e78273. (doi:10.1371/journal.pone.0078273)
148. Steedman M. 2002 Connectionist and symbolic representations of language. In *Encyclopedia of cognitive science*. Nature Publishing Group, Macmillan.
149. Manning CD. 2003 Probabilistic syntax. In *Probabilistic linguistics* (eds R Bod, J Hay, S Jannedy), pp. 289–341. Cambridge, MA: MIT Press.
150. Elman JL. 1990 Finding structure in time. *Cogn. Sci.* **14**, 179–211. (doi:10.1207/s15516709cog1402_1)
151. Pollack JB. 1990 Recursive distributed representations. *Artif. Intell.* **46**, 77–105. (doi:10.1016/0004-3702(90)90005-K)
152. Mikolov M, Burget L, Cernock J, Khudan-pur S. 2010 Recurrent neural network based language model. In *Interspeech, 2010. 11th Annu. Conf. International Speech Communication Association* (eds T Kobayashi, K Hirose, S Nakamura), pp. 1045–1048.
153. Socher R, Manning CD, Ng AY. 2010 Learning continuous phrase representations and syntactic parsing with recursive neural networks. In *Proc. NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop* (eds JD Lafferty, CKI Williams, J Shawe-Taylor, RS Zemel, A Culotta).
154. Socher R, Bauer J, Manning CD, Ng AY. 2013 Parsing with compositional vector grammars. In *Proc. ACL Conf.*, pp. 455–465. Sofia, Bulgaria: Association for Computational Linguistics..
155. Le P, Zuidema W. 2014 The inside–outside recursive neural network model for dependency parsing. In *Proc. EMNLP14*, pp. 729–739. Stroudsburg, PA: Association for Computational Linguistics.
156. Goller C, Kuechler A. 1996 Learning task-dependent distributed representations by backpropagation through structure. In *Proc. Int. Conf. Neural Networks*, pp. 347–352. Washington, DC: IEEE.
157. Rodriguez P. 2001 Simple recurrent networks learn context-free and context-sensitive languages by counting. *Neural Comput.* **13**, 2093–2118. (doi:10.1162/089976601750399326)
158. Fitch WT, Hauser MD. 2004 Computational constraints on syntactic processing in a nonhuman primate. *Science* **303**, 377–380. (doi:10.1126/science.1089401)
159. Fitch W, Martins MD. 2014 Hierarchical processing in music, language, and action: Lashley revisited. *Ann. N.Y. Acad. Sci.* **1316**, 87–104. (doi:10.1111/nyas.12406)
160. Dahlhaus C. 1991 *The idea of absolute music*. Chicago, IL: University of Chicago Press.
161. Bohlman PV. 2002 *World music: a very short introduction*. Oxford, UK: Oxford University Press.
162. Pinker S, Jackendoff R. 2005 The faculty of language: what's special about it? *Cognition* **95**, 201–236. (doi:10.1016/j.cognition.2004.08.004)
163. Jackendoff R. 2009 Parallels and nonparallels between language and music. *Music Percept.* **26**, 195–204. (doi:10.1525/mp.2009.26.3.195)
164. Clayton M. 2009 The social and personal functions of music in cross-cultural perspective. In *The Oxford handbook of music psychology*, pp. 35–44. Oxford, UK: Oxford University Press.
165. Cross I. 2005 Music and meaning, ambiguity and evolution. In *Musical communication* (eds D Miell, R MacDonald, D Hargreaves), pp. 27–43. Oxford, UK: Oxford University Press.
166. Juslin PN. 2013 From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life Rev.* **10**, 235–266. (doi:10.1016/j.plrev.2013.05.008)
167. Hanslick E. 1896 *Vom Musikalisch-Schönen*. Barth.
168. Zangwill N. 2004 Against emotion: Hanslick was right about music. *Br. J. Aesthet.* **44**, 29–43. (doi:10.1093/bjaesthetics/44.1.29)
169. Rebuschat P, Rohrmeier M, Cross I, Hawkins JA. 2012 *Language and music as cognitive systems*. Oxford, UK: Oxford University Press.
170. Cross I. 2012 Music as social and cognitive process. In *Language and music as cognitive systems* (eds P Rebuschat, M Rohrmeier, I Cross, J Hawkins). Oxford, UK: Oxford University Press.
171. Chomksy N. 1980 *Rules and representations*. New York, NY: Columbia University Press.