Implicit learning and recursion

Rohrmeier, M.[1], Dienes, Z.[2], Guo, X.[3], & Fu, Q.[4]

1 Department of Linguistics & Philosophy, Massachusetts Institute of Technology
2 University of Sussex, Brighton, United Kingdom
3 Shanghai Key Laboratory of Magnetic Resonance and
School of Psychology and Cognitive Science, East China Normal University, Shanghai, China
4 Institute of Psychology, Chinese Academy of Sciences, Beijing, China

To appear in:
Francis Lowenthal & Laurent Lefebvre (Eds), *Language and Recursion*, Springer Verlag.

Introduction

Implicit learning constitutes a remarkable human ability to acquire intuitive or unconscious knowledge of environmental structures (Dienes, 2012; Reber, 1989; on the knowledge being unconscious, contrast Shanks, 2005; for a somewhat intermediate position, Cleeremans, 2006). Implicit knowledge governs a large variety of everyday activities ranging from sports to music to language. For example, in the realm of perceptual motor skills, people appear to be able to learn to calibrate golf swings using information that is not conscious (Masters, Maxwell, & Eves, 2009); and can catch balls while consciously holding false theories about how they do this (Reed, McLeod & Dienes, 2010). Native speakers are competent in the use of their language, yet cannot make explicit all the regularities they are actively following. The same holds for music: Most people are competent listeners and perceive musical structure adequately (Koelsch et al., 2000; Bigand & Poulin-Charronat, 2006; Honing, 2011; Deliège et al, 1996), but cannot articulate underlying musical rules without education. Implicit knowledge is assumed to constitute the foundation of music or language cognition, being acquired incidentally from mere exposure and little explicit teaching (Williams, 2009; Rebuschat, 2008; Rohrmeier & Rebuschat, 2012). Accordingly (first or second) language and music acquisition are prime instances of implicit learning (Rebuschat & Williams, 2011; Loui, 2011; Rohrmeier, 2010).

Most implicit or statistical learning studies, however, rely on established methodologies which in turn have implications for the complexity of the structures they use: They employ structures which are based on regular (finite-state) grammars or syllable or tone “words” (i.e. triplets) because they are well-understood and have controlled properties suitable for experimental studies. However, in the ecological cases of language and music, the complexity of the underlying principles of structure building scarcely involves such artificial structures (e.g. “tone words”) and exceeds Markov or finite-state models (cf. Rohrmeier & Rebuschat, 2012; Rohrmeier, Fu & Dienes, 2012). In particular, linguistic as well as musical syntax is governed by tree structures and underlying recursive structures or grammars as part of their principles of structure building (e.g. Chomsky, 1956; Chomsky, 1995; Hauser, Chomsky & Fitch, 2002; Rohrmeier, 2007, 2011; Lerdahl & Jackendoff, 1983; Steedman, 1984; Hofstadter, 1979; Jackendoff & Lerdahl, 2006; Jackendoff, 2011). Hence there is a theoretical gap between the ecological structure of the domains and the structures by which the implicit acquisition is studied in the domain. Accordingly one core aspect of this gap concerns the concept of *recursion* as a core principle of structure building in language and music. Thus given that language and music employ recursive structures, the underlying learning and processing mechanisms have to be capable of dealing with structure that can be recursively generated. Before considering its empirical implications, we discuss the concept of recursion in the context of music and language.

The concept of recursion in artificial grammars

The relationship between the concepts of recursion and related debates concerning formal languages, the Chomsky hierarchy (and, particularly, non-finite-stateness) is not straightforward (cf. Fitch & Friederici, 2012; Tomalin, 2007; Martins, 2012; Lobina, 2011; for an account of implicit learning and recursion see Rohrmeier, Dienes, Guo & Fu, 2014) In this section we review the concept of recursion in formal languages – grammars as well as processes dealing with them – in order to provide a basic, somewhat less formal introduction to readers with different backgrounds and unfamiliar with the concepts and subsequently to relate these concepts to the empirical instantiations in music and language as well as the field of implicit learning.

Recursion serves as a construction principle for infinite sets. Grammars are sets of formal rules to describe a finite or infinite set of sequences (a *language*) through construction principles. While a finite number of finite sequences could be simply listed in a set, an infinite number of sequences (or an infinite sequence) requires an indirect definition to describe them as a set. The (mathematical) definition of an infinite set by a finite set of production rules is grounded in recursion (by virtue of the multiple application of some production rules). Generally, an account of recursion is relatively straightforward:

> (i) A definition of a structure is recursive, if it includes itself as part of the definition.

A common mathematical example employing this definition is the Fibonacci sequence (0, 1, 1, 2, 3, 5, 8, 13, …) in which each term is defined as being the sum of the previous two terms (given 0,1 starts the sequence), i.e. fib(n) = fib(n-1) + fib(n-2) with fib(1)=0, fib(2)=1; note that the definition of the starting condition is crucial for the recursion to terminate. Definition (i) does not directly relate to formal grammars and further does not directly translate into a difference between finite-state and context-free grammars. When transferred to formal languages, the definition may state:

> (ii) A formal language is recursive, if a sequence of productions leads to the production of a non-terminal symbol (as part of the string) that was part of its earlier production.

Terminal symbols are the elements that a string (i.e. sequence) is actually composed of; for example, words in a sentence, or letters in the letter strings often used in artificial grammar learning experiments. Non-terminal symbols are variables that are used as part of a rewrite process in order to produce a final sequence of terminal symbols. For example, consider the following production (rewrite) rules[1]*:*

[1] Here we use the convention in the artificial grammar learning literature in experimental psychology, with numbers referring to non-terminals and capital letters as terminals; another convention is to use capital letters for non-terminals and lower case letters for terminals. The formal

1. [0] -> M [1]
2. [1] -> T [1]
3. [1] -> Q [2]
4. [2] -> B [1]
5. [1] -> Q

The letters M, T, Q, and B are the terminals; the symbols [0], [1] and [2] are non-terminals. The first rule uses the starting symbol [0] to rewrite it according to the right-hand side as M and the nonterminal [1]. In other words, M can start a string. So [0] becomes M [1]. We still need to rewrite further (because the generation cannot terminate before there are no nonterminals in the string), which we can do by rewriting it as T and itself (rule 2). So M [1] becomes MT [1]. Rule 2 is thus recursive, by definition (ii). Rules 3 and 4 together are also recursive (though not one directly visible in either rule alone) because the further production of the nonterminal [1] ultimately yields another instance of itself. Rule 5 produces the single symbol Q, which ends the production sequence.

Rewrite rules generate a set of strings (in our case, e.g. MQ, MTQ, MTTQ, MQBQ, …). The processing of such strings (in order to decide whether a string is part of the grammar or not) requires an *automaton.* The figure below constitutes a representation of the *finite-state automaton* (or *finite-state machine*) corresponding to the grammar above; it is the formal device to recognise strings that are generated by the rules above. Hence the rules are indirectly represented in this graph and it also expresses all production sequences that the rules can possibly generate by different paths. Because regular grammars as well as finite-state automata provide a formally equivalent characterisation of regular languages, finite-state automata are frequently used as representations of the underlying grammar and sequence structure in the artificial grammar learning literature (i.e. rewrite rules 1-5 are represented by the finite-state automaton; note that non-terminals map to states and terminals to arrows).

---

definition of a formal grammar is the 4-tuple G = (*N, Σ, P, S*), for which *N* represents the set of non-terminal symbols, Σ the set of terminal symbols (surface symbols), *P* the set of generative production rules and *S* the starting symbol (from N).

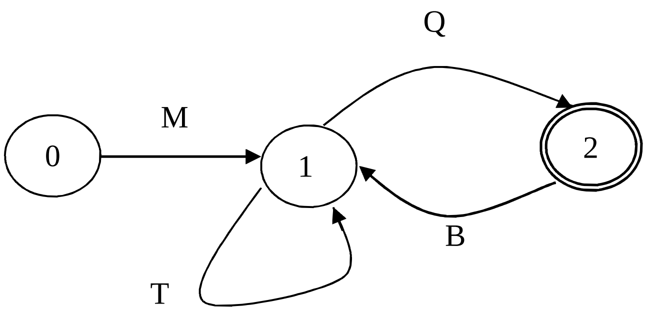

Fig. 1. The finite-state automaton that corresponds to the regular grammar defined above. The states are represented as circles and transitions between states as arrows. The starting state is [0]. The final (accepting) state [2] is denoted by a double-circle.

Regular languages (finite-state grammars) are defined by using rewrite rules where terminals on the right-hand side of the production are rewritten to only one side of a nonterminal (as in the above rules where terminals were always written to the left of the nonterminal, defining a *right-linear regular grammar*). As the example illustrates, regular languages may involve recursive rules, specifically *tail recursion*, i.e. a form of recursion that produces a sequence that recursively expands to one side of the sequence (left or right; note, however, that the combination of both right and left-regular production sequences instantiates a grammar that is no longer regular). Such instances of tail-recursion correspond to cycles in the finite-state automaton.

A context-free grammar allows for rules rewriting nonterminals to any sequence of terminals and nonterminals on the right-hand side. This leads to a whole different form of expressive power including *centre-embedding recursion* which results in a sequence with self-embedding within a sequence. A famous, most simple example of this is the language A^nB^n (i.e.: after any number of As, have the same number of Bs) created from the rules:

6. [S] -> A [S] B
7. [S] -> ε

(or the rules [1] -> A [2], [2] -> [1] B, [1] -> ε, if rules are restricted to binary productions).

Applying rule 6 once produces A[S]B, twice produces AA[S]BB, three times AAA[S]BBB, then applying rule 7 produces the string AAABBB. Obviously rule 7 can be applied after any number of applications of rule 6 (note that this leads to a tree representation in which the symbols produced by each rule application are the children of the node corresponding to the left-hand side variable (in the present example [S]) that is rewritten by the rule; in the context of our example it results in a ternary tree in which each new subtree is embedded in between an A and B symbol). As rewriting occurs either side of the non-terminal, these rules cannot be represented by a finite state diagram. For example, the following (recursive) finite state diagrams (Fig. 2) do not produce all A^nB^n strings and also produce strings which are not A^nB^n : (see Fig. 3 for a better matching representation of A^nB^n for finite *n*).

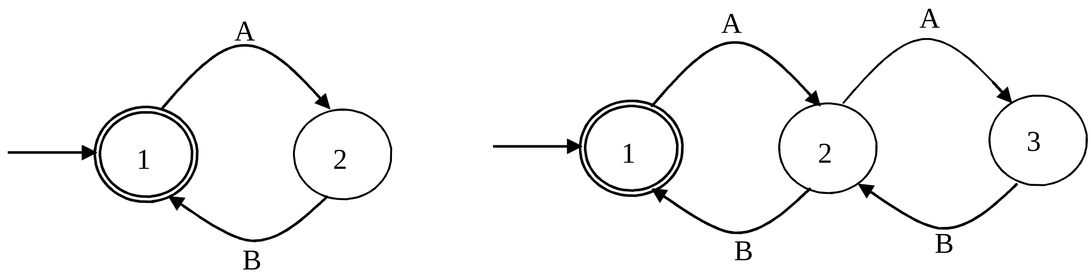

Fig. 2. Finite-state automata for different languages over {A,B}.

Another very simple example of a recursive context-free language would be the *Dyck language,* which describes expressions of all correct nested bracketings (such as [ ], [[ ] [ ]], or, [[[][[][]][]][[]]]), which is empirically relevant e.g. with respect to grouping structure in music (Lerdahl & Jackendoff, 1983; Bod, 2002). It is important to note here that the A^nB^n or Dyck language are two of the *most simple* examples of context-free languages that can be constructed, which feature a structure of recursively nested brackets that is typical for context-free languages. However, they are not representative of the potential expressiveness and complexity of context-free languages (the grammar of a programming language or (simplified) English may be expressed with context-free languages).

Supraregular (context-free and higher) recursive grammars also produce various forms of mirror symmetries (Hopcroft, Motwani & Ullman, 2000). While mirror symmetries exhibit properties that are easily modelled by variants of the A^nB^n grammar, translational symmetry such as $A_1A_2A_3B_1B_2B_3$ requires (mildly) context-sensitive grammars. Because of these interesting properties, sometimes debates restrict the use of “recursion” to languages above regular languages.

As the construction rules of such cross-serial dependencies or translational symmetries are rarely made explicit, we would like to show them:

8. $S \rightarrow a_i\ S\ T_i$
9. $S \rightarrow \varepsilon$
10. $a_i\ T_j \rightarrow a_i\ b_j$
11. $b_j\ T_i \rightarrow T_i\ b_j$

The grammar first constructs (a T) pairs and then employs the context-sensitive rules 10 and 11 to create b symbols from T and swap as and bs until they are in the right order. (Note that a generation by this grammar cannot described by a tree; further note again that the cross-serial dependency grammar is a very simple and not representative case of the expressive power underlying context-sensitive grammars)

The parsimony argument

Structures generated by recursive grammars (be they regular, context-free or of higher complexity) may exhibit particular features that cannot be expressed by simpler models. The self-embedding/centre-embedding nature of context-free grammars entails nonlocal dependencies. While regular grammars can express some forms of potentially infinite long-distance dependencies (see Chomsky, 1956; Cleeremans, 1993; Rohrmeier, Fu & Dienes, 2012), the structures linking both ends of several dependencies have to be represented multiply – and nested nonlocal dependencies hence require a factorially growing form of representation. Particularly with respect to the latter, context-free grammars constitute a considerably simpler and more parsimonious form of representation (and model informing production/perception). We refer to this as the *Parsimony Argument*. Consider for example representing A^nB^n with a finite state grammar. It can be done for a finite n. The grammar on the left works for n = 1; the (non-recursive) grammar on the right works for n = 2:

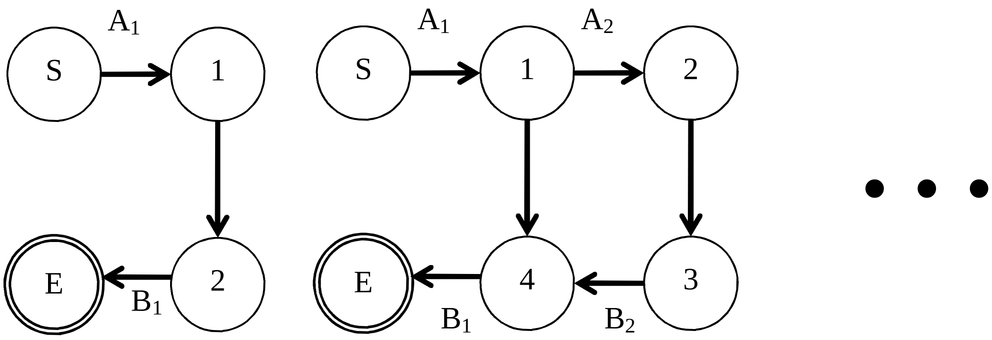

Fig. 3. Finite-state automaton approximations of the A^nB^n language for finite numbers of n. The displayed automata can be extended to any number n by adding 2n additional states.

But for each increase in n, another set of states have to be added to the representation. If people were trained to detect A^nB^n for n= 1, 2 and generalised to higher n's, it may be more parsimonious to postulate the representation of a context-free grammar that contains only two rules (or a corresponding processing mechanism) than of a finite state machine where the subject constructed more, potentially an factorially growing number of, states or rules than needed to account for what they were exposed to. The regular grammar representation cannot make elegant use of recursion to express the repeating regularity underlying the structure. In addition, if such a catch-all finite-state model postulates learning states or chunks not required for learning the training material to predict participants' generalisation performance, the explanation is similarly less parsimonious.

Altogether that is, postulating the mental representation of a grammar or corresponding processing device (automaton) lower down the Chomsky hierarchy, just because one can, is not necessarily to give the simplest theory in particular when it comes to context-free structures in the case of ecological language acquisition that are far more complex than the A^nB^n language. Postulating a representation of higher Chomsky hierarchy complexity buys less description length and better representation for the price of what may be little more than the use of a specific form of memory buffer (such as a stack or a queue).

Recursive context-free structures are frequently represented with hierarchical tree-based representations. These hierarchical representations involve recursion in a way such that a subtree may recur as a substructure at another level of the tree (even in the case of tail-recursion, i.e. regular tail-recursion constitutes a tree that is degenerated to a list). For instance, relative clauses in language or modulation in music are examples of such recursive structures consisting of linguistic or musical phrases embedded within phrases. However, hierarchical tree-based representation does not in itself imply that the structure is recursive. There are meaningful hierarchical representations of music (e.g. piece – section – phrase – motif) or for instance some kinds of phonological trees that are hierarchical but may not be recursive.

Recursion in language and music

There are many examples of linguistic and musical structures that exhibit tail-recursion and centre-embedding recursion as their construction principles. For instance, the generative syntax model of tonal harmony (GSM; Rohrmeier, 2011) provides a concise specification which features of musical harmonic sequences exhibit recursion: applied secondary or diatonic dominants are instances of tail recursion, while centre-embedding recursion is created by modulation (remarkably, first pointed out by Hofstadter, 1979; cf. Giblin, 2008) and functional region expansion, both creating the potential for nested non-local dependencies (see Rohrmeier, 2011, for details). Further, musical grouping structure involves recursive nesting (Lerdahl & Jackendoff, 1983; Bod, 2002). Similarly it is easy to come up with two examples of tail-recursion and recursive centre-embedding in language (see 1a, 1b and 2 below). Comparably, Steedman (2002) and Jackendoff (2009, 2011) argued similar cases for embedded complex action or movement sequences.

(1a) This sentence continues and continues and continues and continues …
(1b) The green, red, yellow, purple, … balls are in the box.
(2) Peter who gave Anne who helped Frank who … paint the fence the ball laughed.

Implications regarding mental representation, learning, processing and methodologies

Based on this theoretical background the empirical investigation of recursion in learning and processing involves a number of methodological and philosophical issues. The core point revolves around the issue of finiteness: empirical research deals with finite materials, yet as outlined above, while recursive rules or mechanisms are a means for the description or production of (potentially) infinite sets of sequences, finite sets *potentially* can be acquired, processed and represented by lists of examples or chunks, simpler forms of grammars, etc.. Hence though it is theoretically difficult to disprove simpler forms of representation, a number of methodological steps may assure to render alternative explanations much less compelling or plausible.

As remarked by Dulany, Carlson, and Dewey (1984), Reber (1993) and others, the structure according to which the materials are constructed need not be equivalent to the structures that are mentally represented. The reflection of this insight has shaped the prevalent notion in artificial grammar learning research that rarely the regular grammars themselves are assumed to be acquired or represented (as rules or finite-state automata; see e.g. Pothos, 2007). With respect to more complex recursive structures, this issue is more intricate. First, the view that chunk or n-gram learning suffices to account for the structures in natural language is theoretically inadequate (see e.g. Chomsky, 1956). Second, a positive finding of acquired recursive structures has implications for the underlying cognitive architecture: the mental representation may require not only a representation that is isomorphic to one dealing with the structure (such as a tree), but also an additional parsing process dealing with these structures (including potential mechanisms of revision and backtracking to solve ambiguities and gardenpath sequences). For instance, the implicit learning of the *Dyck language* (of which A^nB^n is the simplest subset and example), will require the instantiation of a parsing process that keeps track of each open bracket and "ticks off" the innermost open bracket (and potentially triggers another related action) once a closing bracket is encountered. A sequence is perceived as ungrammatical when one closing bracket on an empty stack or an open bracket at the end of the sequence are encountered. This example illustrates that this process does not necessarily involve the mental formation of a representation of a tree structure from the encountered sequence: merely the parsing process itself suffices to deal with the structures. It further illustrates that the involved parsing mechanism may but need not be recursive itself: the parsing process of the context-free, recursive bracketing structure, merely involves memory (e.g. a stack) and an *iterative* (potentially nonrecursive, or not explicitly recursive) process[2] that adds or ticks off brackets. Generally, processing (and dealing with) recursive structures is potentially achieved without recursive representations or recursive mechanisms. With respect to the learning process, it is an implausible assumption that the parsing mechanism is acquired together with its examples. But it may be the case, that an established cognitive mechanism to deal with embedded structures is triggered as the most efficient to deal with the perceived sequences.

---

[2] Note that the distinction between recursive and iterative algorithms involves a notion of recursion that avoids encompassing all kinds of iteration (cf. Tomalin, 2007).

Furthermore, a separation of training sequences and testing sequences such that testing sequence are novel and force generalisation beyond exemplar and chunk memorisation (cf. Rohrmeier, Rebuschat & Cross, 2011; Rohrmeier & Cross, 2010) warrant that a more general and abstract representation is acquired. A second issue concerns the examined structure from which materials are taken. Some grammars that may be expressed by formal rules like context-free rules may not exhibit recursion (such as Saffran, 2001, 2002, and the *Brocanto* language, e.g. in Opitz & Friederici, 2003).

There are also two issues concerning the interpretation of the findings and their generalisation: Does the finding of learning of the well-studied A^nB^n language indeed warrant the conclusion that *generally* (all or most) context-free grammars can be learnt (implicitly)? Whether such a finding expands to general cases of context-free languages (closer to ecological sequences) or languages of the complexity of formal languages like PASCAL or LISP remains a barely addressed issue. Second, despite its theoretical importance, the Chomsky hierarchy (Chomsky, 1959; Chomsky & Schützenberger, 1963; Jäger & Rogers, 2012) does not constitute the only way to carve out the infinite space of sequential structures. It may well be that other distinctions deriving from processing issues (the types of sequences that can be processed by a sequential neural network, like the SRN, or a graphical model [like an HMM]) may provide implementation-motivated ways (cf. Marr, 1982) to distinguish between types of structures based on the way in which brain mechanisms are able to deal with them. These distinctions may not coincide with the Chomsky hierarchy.[3]

Implications for mental representation or processing of recursive structures

As argued above, mechanisms of processing or parsing of recursive structures may not require (but could employ) recursion itself. In fact, the ability to deal with recursive structures may not require a representation of the recursive rules or the full recursive structure: e.g. instead of a full tree, a partial mental representation may potentially suffice, for instance, for music, without having the full structure ever represented at once (see also Rodrigues, Wiles & Elman, 1999, for an account of how a SRN model could learn A^nB^n structure). However, when materials are designed in ways such that chunking or similar accounts are insufficient (see for instance the modelling of Dienes & Longuet-Higgins, 2004, in Rohrmeier, 2010; or Koelsch, Rohrmeier, Jentschke & Torrecuso, 2013), findings reveal that at least one of the potential mechanisms to deal with nonregular structures is at work. While finite-state machines offer ways to deal with finite numbers of embeddings, the parsimony argument may act in favour of the learning / representation of non-finite state machines: mechanisms isomorphic with recursive representation or

---

[3] Note also, that the distinction between context-free and context-sensitive structures is, in practice, separated by worlds of complexity which are not adequately reflected by the simple distinction of centre-embedded and (mildly context-sensitive) cross-serial structures, e.g. A1A2A3B3B2B1 and A1A2A3B1B2B3.

processing are less complex than all-encompassing catch-all finite-state machines. In this context, computer simulations with exemplar or chunking learning mechanisms provide the most powerful way to establish that certain structures cannot be acquired by a plausible chunk learning process. Further it is important to note that the frequent counter argument, that the simplicity of recursive rules is traded for remarkable complexity involving recursion rather than no recursion in the processor, does not hold here since processing mechanisms powerful enough to deal with recursion (e.g. empirically occurring levels of centre-embedding) may not involve recursion themselves. Finally, however, such accounts will have to be based on Bayesian model comparison formally taking into account the model complexity involved (compare, e.g., Perfors, Tenenbaum & Regier, 2011). The *Parsimony Argument*, however, suggests that recursive models can be favoured despite the apparent "simplicity" of using processors lower down the Chomsky hierarchy. If one has to postulate that e.g. a chunking processor acquires chunks it was not exposed to in order to correctly classify in the test phase, simplicity favours a model more powerful than a chunker.

Evidence from empirical research

AGL in language and music

When Reber (1967) coined the term "implicit learning" by looking at how people learn artificial grammars (following George Miller explorations with artificial grammar learning; Miller, 1967), he decided to start at the bottom of the Chomsky hierarchy, with regular grammars. The choice was a sensible starting point. Regular grammars are still sufficiently complex that people do not consciously notice all of the structure, while providing structure that can be unconsciously learnt. Regular grammars have provided decades of work on clarifying some simple things people can learn (see Pothos, 2007 for a review). For example, when exposed to a regular grammar, people can learn chunks of especially two or three successive elements (sometimes four) (e.g. Servan-Schreiber & Anderson, 1990; Perruchet & Pacteau, 1991; Rohrmeier, Rebuschat & Cross, 2011; Rohrmeier & Cross, 2010, 2013); people can learn specific training exemplars and use these as analogies (e.g. Brooks, 1978; Jamieson & Mewhort, 2009); and people can learn specific repetition structures (Brooks & Vokey, 1991). A repetition structure is a way of learning a long distance relationship through memorisation (and thus is important to control in work on learning non-regular grammars.) The strings of letters MTTVT and ABBCB have the same repetition structure, which could be represented as 12232, meaning the first letter is unique, the second letter is different from the first, but the third letter is the same as the second, the fourth letter is unique and the final letter is the same as the second. Chunks, exemplars and repetition patterns can be learned implicitly (e.g. Scott & Dienes, 2008). While these structures fall short of learning all structure in any regular grammar, the question is still left open of just what structure in non-regular grammars can be implicitly learned. Given the apparent qualitative difference between implicit and explicit learning (Dienes, 2012), our main interest is in what structures can be implicitly learned (note that explicit learning of context-free grammars such as a programming language comes as no surprise).

Learning of context-free grammars in language and music

As outlined above, natural languages and musical structures exhibit some structural principles that require (at least) context-free grammars to be expressed. There are a number of studies that explore learning of A^nB^n and related structures (e.g. Fitch & Hauser, 2004; Perruchet & Rey, 2005; Friederici et al, 2006; Vries et al, 2008; Bahlmann et al, 2008; Hochmann et al, 2008; Lai & Poletiek, 2010; Poletiek, 2011; Uddén et al, 2012). There have been several reviews discussing these studies in the context of recursive processing and learning (e.g. Fitch & Friederici, 2012; Rohrmeier, Fu & Dienes, 2012). Accordingly here we focus on recent work that goes beyond the scope of these studies and explored implicit learning of more complex context-free grammars that were modelled to reflect ecological features of music and language more closely.

Rohrmeier, Fu & Dienes (2012) constructed materials that modelled abstract word order in linguistic relative clauses, such as "the man, who met the woman, sold the chair". To construct a simplified model language, the artificial grammar accordingly features three categories *N, V, R* representing categories nouns, transitive or intransitive verbs, and a relative pronoun. Each of the categories *N* and *V* featured four possible words, whereas the *R* class consisted only of one word. For modelling the embedded structure, two core distinctions were taken into account: the embedding structure could be either left- or right-branching; and structures could be either centre-embedded or tail-recursive. All four possible combinations were explored each with one group in the study. Starting from the main clause "N V" or "N V N" (modelling a simple sequence such as "boy kisses girl"), relative clauses of the form "R N V" or "V N R" (centre-embedding and right/left-branching) or "R V N" or "N V R" (tail-embedding and right/left-branching) could be attached to each noun with up to two levels of embedding. Further relative clauses without noun classes were used: "R V" or "V R" (right and left-branching). Example sequences are: "NV(VNR)N", "N(RN(RNV)V)VN" for left-branching and right-branching embedding of the first and second order. Stimuli were rendered as spoken sequences. The experiment used four experimental and control groups which all first completed a learning phase being exposed to 168 examples from the grammar featuring up 0 up to 2 levels of embedding (the control group was trained on random sequences). The experimental phase employed the Process Dissociation Procedure (Jacoby, 1991) presenting pairs of grammatical stimuli that were presented in the training (old-grammatical) or only in testing (new grammatical) and ungrammatical stimuli that either featured systematic violations of one embedded subsequence or a random sequence. Finally participants completed a category test in which participants had to choose two out of three words that would belong to the same category.

Results showed that people could classify both new and old grammatical structures above chance with little difference between them. A comparison between responses to random and layer-violating ungrammatical sequences further indicated that participants had acquired detailed knowledge about subtle difference in the grammatical structure. The category test revealed that participants acquired implicit knowledge of word classes.

In greater detail, analyses revealed that participants were performing better for sequences with violations in the first or second layer than the third layer suggesting that higher-order embeddings turned out more demanding. Finally participants outperformed controls on local as well as nonlocal violations. A subsequent logistic regression analysis was carried out to examine the extent to which responses could be explained by local n-gram learning or repetition structure. It was found that nonlocal dependencies formed the greatest predictor of responses although n-grams and local repetition structure had a small yet significant impact. Crucially, subjects were sensitive to long distance relations above and beyond their sensitivity to n-grams or repetition patterns, at either the level of terminals or non-terminals. Further, people were sensitive to the long-distance dependencies after partially out sensitivity to fixed distance dependencies, consistent with the knowledge of long distance dependencies being structure sensitive. Accordingly the findings show evidence for implicit learning of complex (nonlocal and local) dependencies as constructed by a recursive context-free grammar. Finally, the results suggested that tail-recursive structures were better learned than centre-embedding structures and that left-branching was better learned than right-branching when the grammar was tail-embedding. This performance difference correlates with Hawkins's (2004) performance and correspondence hypothesis for natural languages – supporting the claim that the more complex type of grammar may impede performance and therefore be less frequently established across language varieties.

To investigate implicit learning of musical harmonic structure (which is recursive and context-free; see Rohrmeier, 2011, and Steedman, 1996), Rohrmeier & Cross (2009) as well as Rohrmeier (2010, ch.4) used the same experimental and structural design as Rohrmeier, Fu & Dienes, transferring the entire setting to the musical domain. Instead of three word classes three chord classes were constructed using octatonic scales, two of which contained 4 chords and one contained a single dissonant chord. Octatonic scales were used to make it possible to employ familiar categorisable materials (major and minor chords) in a novel, unfamiliar context (not a major or minor scale). Since musical structure is organised predominantly with left-branching (i.e. supporting *goal-directed* sequences), only tail- and centre-embedding left-branching structures were employed. In analogy to the linguistic findings, results indicated that Western musicians acquired generalised implicit knowledge of the grammatical sequences, in particular local as well as nonlocal structures. According to the logistic regression analysis, n-gram learning was only a weak to non-existent predictor (exp. 2) whereas grammatical structure turned out to be the strongest predictor. Again, centre-embedding materials were less well learned than the sequences featuring tail-embedding. This suggests that there may be a domain-general processing / performance advantage for simpler tail-recursive sequences to centre-embedding structures in line with Hawkins's (2004) theory.

In further experiments, the same musical paradigm was employed to test implicit learning in Chinese musician and nonmusician subjects. The results showed that all groups acquired both grammars above chance, however, Western musicians outperformed Chinese musicians significantly. Further the difference between Chinese musicians and nonmusicians was comparably small (only with respect to random and layer 1 violations). One potential reason why the structures were harder for Chinese participants may be an effect of acquired implicit musical knowledge (cf. Rohrmeier & Cross, 2011, 2013), given that traditional Chinese music features a stronger emphasis on melodic structure and pentatonic scales (even in pop music) and less on harmonic structure.

Learning context-sensitive grammars in music, poetry and movement

Symmetry is a structure that can be generated recursively, and requires at least a context free grammar to parse. Symmetry occurs not only in language (e.g. centre embedding of the form $A_1A_2A_3B_3B_2B_1$ and cross serial dependencies of the form $A_1A_2A_3B_1B_2B_3$) but also in art, including in music. For example, serialist (twelve tone) music makes use of the symmetries of inversion (produced by changing upward to downward motion and vice-versa) and retrograde (produced by placing a mirror vertically beside a line of music score, i.e. generating a reversed “retrograde” version of a melodic line). Dienes and Longuet-Higgins (2004) provided preliminary evidence that experts in serialist music could implicitly learn to become sensitive to a particular musical symmetry (inverse, retrograde, inverse retrograde and transpose (i.e. copy)) after brief exposure to it. Kuhn and Dienes (2005) simplified the materials and showed that musically unselected participants after exposure to musical inversions, increased their liking of inversions compared to non-inversions compared to an untrained control group. Further, people were unable to classify which strings were rule following and which were not, demonstrating the implicit nature of the knowledge that expressed itself in liking. Kuhn and Dienes controlled chunk strength in the materials to rule out this simple explanation. However, for the actual materials used, which were of fixed length, Kuhn and Dienes (2008) showed an SRN could learn a set of long distance associations between a tone in one position and a tone a fixed distance later. That is, subjects need not have learnt the symmetry per se in order to show learning on the test set.

The music inversion results have been further explored using Chinese Tang Poetry. Chinese words are spoken in one of four tones, which are traditionally categorized into two groups, ping and ze. Treating ping and ze as opposites, Jiang et al (2012) constructed artificial poetry in which successive lines of each poem were ping-ze inverses of each other and hence constitute instantiations of cross-serial dependencies. Jiang et al strictly controlled both chunk strength and repetition patterns (Vokey and Brooks, 1994). People asked to memorise such poems later classified new poems (of the same line length) as better formed when they showed the inverse pattern. Further, on a trial by trial basis, they almost exclusively claimed to classify based on pure guessing or on intuitive feelings they could not explicate further. That is, the learning was implicit. Li et al (2013) showed that poems expressing a retrograde symmetry could also be classified as well formed following training on retrogrades (once again controlling chunk strength and repetition pattern). Retrogrades can be parsed using a mechanism that can deal with a context free grammar but inverses require something more than context free. So which is easier – retrogrades or inverses? Li et al showed that implicitly learning retrogrades was harder than implicitly learning inverses. That is, grammatical complexity as expressed by Chomsky hierarchy predicts the order of difficulty the opposite way (given it is actually retrogrades and inverses that subjects are learning in the poetry experiments). Both inverses and retrogrades require a buffer memory mechanism to be parsed; inverses could be most simply parsed using a first in- first out buffer (a queue) and retrogrades using a first in-last out buffer (a stack). Thus, the results are consistent with implicit learning using a memory functionally like a first in-first out buffer for non-regular grammars. Jiang (2010) showed an SRN could also learn the poetry inversion materials. She also showed that people only learnt the inversion when ping and ze categories were used; when a linguistically arbitrary binary category was used for the tones, people did not learn (thus people were learning a relation over tone classes and not tones). Similarly the SRN only learns when given the ping ze categories as input. Guo et al. (in preparation) also showed the SRN could learn the retrograde materials but less well than the inversion materials, just like people. The match between the characteristic behaviour of people and SRN is remarkable. But have either actually learned the symmetry per se?

Guo et al. performed a crucial experiment looking at transfer to different lengths. If people have simply learnt a long distance association, there is no reason for the knowledge to transfer to different length test items. However, when people were trained on inversions where the length of each line was five words, they could generalise to poems of both four words and six words. There was a notable performance decrement from trained to untrained lengths. Thus, the simplest explanation may be that people form a graded fixed length association (for the ability of people to learn fixed length associations see Remillard, 2008). That is, it is not yet clear people could treat length itself as a variable. If when trained on different lengths, people could generalise to yet new (longer) lengths with facility, it would motivate the claim that people could learn to map a vector of unspecified length to its negation, i.e. learnt the inverse rule per se. In the meantime, the simplest explanation is that people learnt fixed long distance associations in a graded way. However, people may also have a partial grasp of a relevant context-sensitive grammar (though presumably they have not explicitly or symbolically represented the context sensitive rules we gave above, but might instantiate a process and memory representation to capture such cross-serial dependencies) – people's natural, conscious, yet automatic, appreciation of symmetry in nature, art, faces and scientific theories (Dienes et al, 2011; Westphal-Fitch et al, 2012) motivates our conjecture that people can detect mirror symmetries implicitly as well as explicitly. This is an area for further exploration. Guo et al (in preparation) also found that the SRN characteristically learns to generalise from length five to lengths four and six for these materials, in a similar way as people. As for people, there was a notable decrement in performance from the trained length to the neighbouring lengths. Thus, so far it remains fair to characterise the SRN as a graded finite state machine for our materials (as Cleeremans, 1993, did). But we seek to test this claim and see if the SRN will become a graded recursive structure processor (when trained on different lengths) (compare Rodrigues et al, 1999, who found the SRN could learn A^nB^n, which gives us optimism).

We have implemented the same inversion structure as used by Kuhn and Dienes (2005) in movement sequences (see Dienes et al 2012). Kuhn and Dienes used the scale of C major for all their "melodies". Thus, each melody could be construed as a movement around a circle (clock arithmetic with 8 points on the clock face). Dienes et al (2012; in preparation) asked subjects to trace a circle with their finger, so in effect the subjects traced out inversions isomorphic to musical ones. The materials were controlled for chunk strength and repetition structure. Subjects came to like the inversions more than non-inversions to a greater extent than untrained controls, even when just considering subjects who claimed no awareness of the symmetry at the end of the experiment. We hope that the movement paradigm will be one that other researchers can take up (more easily than music or Chinese tonal paradigms) to explore in more detail how the implicit system comes to parse recursively generated structures, an under-researched topic in implicit learning.

Conclusion

Although implicit learning is a well-established, long-standing research tradition, there is a gap between the types of structures used implicit learning studies and the ecological complexity and type of structure (such as in music and language). In particular, language and music exhibit a variety of recursive features which cannot be modelled by regular languages, but require context-free or (mildly) context-sensitive structures. We discuss empirical issues concerning the exploration and the conclusion drawn from the study of implicit learning of recursive structures. Although finite cases of recursive structures could (theoretically) always be explained by catch-all exemplar-based or finite-state representations, such accounts are challenged by their inefficient representation (the *parsimony argument*) facing recent findings concerning learning nonlocal dependencies and the generalisation of acquired knowledge. We report empirical studies which suggest that musical and linguistic recursive context-free structures can be acquired implicitly in ways which are not explained by n-gram, repetition structure, or fixed length association learning. Another series of studies suggest that simple context-sensitive cross-serial dependencies are acquired implicitly, above and beyond the learning of n-grams and repetition structures (but the case for the knowledge extending beyond fixed long-distance associations is not yet made). Taken altogether there is a growing body of evidence that supraregular structures can be acquired implicitly – findings which considerably extend the present knowledge about the limits of implicit learning and challenge the differences in complexity predicted by the Chomsky hierarchy.

References

Altmann G, Dienes Z, Goode A (1995) On the modality independence of implicitly learned grammatical knowledge. Journal of Experimental Psychology: Learning, Memory, & Cognition 21: 899-912.

Bahlmann J, Schubotz RI, Friederici AD (2008) Hierarchical sequencing engages Broca's area. Neuro Image 42: 525–534.

Bahlmann J, Schubotz RI, Mueller JL, Koester D, Friederici AD (2009) Neural circuits of hierarchical visuo-spatial sequence processing. Brain Research 1298: 161–170.

Bigand E, Perruchet P, Boyer M (1998) Implicit learning of an artificial grammar of musical timbres. Cahiers De Psychologie Cognitive-Current Psychology of Cognition 17(3): 577–600.

Bigand E, Poulin-Charronnat B (2006) Are we "experienced listeners"? a review of the musical capacities that do not depend on formal musical training. Cognition 100: 100–130.

Bod, R. (2002). A Unified Model of Structural Organization in Language and Music. *Journal of Artificial Intelligence Research*, 17(2002): 289-308.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. *Cognition and categorization*, *3*, 170-211.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). Journal of Experimental Psychology: General, 120(3), 316–323.

Chomsky N (1956) Three models for the description of language. *IRE Transactions on Information Theory IT*-2: 113–124.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, *2*(2), 137-167.

Chomsky, Noam; Schützenberger, Marcel P. (1963). The algebraic theory of context free languages. In Braffort, P.; Hirschberg, D.. *Computer Programming and Formal Languages*. Amsterdam: North Holland. pp. 118–161.

Cleeremans, A. (1993). *Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing*. Cambridge, MA: MIT Press

Cleeremans, A. (2006). Conscious and unconscious cognition: A graded, dynamic, perspective. In Q. Jing, M. R. Rosenzweig, G. d'Ydewalle, H. Zhang, H.-C. Chen & K. Zhang (Eds.), *Pro-*

gress in Psychological Science around the world. Vol I. Neural, Cognitive and Developmental Issues. Hove: Psychology Press, pp. 401-418.

Deliège, I., Mélen, M., Stammers, D., & Cross, I. (1996). Musical schemata in real time listening to a piece of music. *Music Perception*, *14*(2), 117–160.

Dienes Z, Altmann G, Kwan L, Goode A (1995) Unconscious knowledge of artificial grammars is applied strategically. Journal of Experimental Psychology: Learning, Memory, & Cognition 21: 1322–1338.

Dienes, Z., Kuhn, G., Guo, X. Y., & Jones, C. (2011). Communicating structure, affect and movement: Commentary on Bharucha, Curtis & Paroo. In Rebuschat, P., Rohrmeier, M., Cross, I., Hawkins (Eds), *Language and Music as Cognitive Systems*. Oxford University Press (pp 156-168).

Dienes, Z. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat & J. Williams (Eds), *Statistical Learning and Language Acquisition*. Mouton de Gruyter Publishers (pp. 337 - 364).

Dienes, Z. & Longuet-Higgins, H. C. (2004). Can musical transformations be implicitly learned? *Cognitive Science, 28, 531-558*.

Dulany, D., Carlson, R., & Dewey, G. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology*, *113*, 541–555.

Fitch W, Hauser M (2004) Computational constraints on syntactic processing in a nonhuman primate. Science 303: 377–380.

Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: an overview. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences*, *367*(1598), 1933–55. doi:10.1098/rstb.2012.0103

Frensch PA, Rünger D (2003) Implicit Learning. Current Directions in Psychological Science 12(1): 13–17.

Friederici AD, Bahlmann J, Heim S, Schubotz RI, Anwander A (2006) The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences,* 103(7): 2458–2463.

Giblin, I. (2008). *Music and the Generative Enterprise. Situating a Generative Theory of Tonal Music in the Cognitive Sciences*. University of New South Wales.

Hauser, M. D., Chomsky, N., & Fitch, T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.

Hawkins, J. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hochmann JR, Azadpour M, Mehler J (2008) Do humans really learn anbn artificial grammars from exemplars? *Cognitive Science* 32: 1021–1036.

Hofstadter, D. R. (1979). *Godel, Escher, Bach: an eternal golden braid*. London: Penguin.

Honing, H. (2011). The illiterate listener. On music cognition, musicality and methodology. Amsterdam: Amsterdam University Press.

Hopcroft, J. E., Motwani, R. & Ullman, J. D. (2000). Introduction to automata theory, languages and computation, 2nd edn. Reading, MA: Addison-Wesley.

Jackendoff, R. (2011). What is the human language faculty?: Two views. *Language*, *87*(3), 586-624.

Jackendoff, R. (2009). Parallels and Nonparallels between Language and Music. *Music Perception*, *26*(3), 195–204. doi:10.1525/mp.2009.26.3.195

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and What's specical about it? *Cognition*, *100*, 33–72.

Jacoby, L. L. (1991). A process Dissociation Framework: Separating Automatic from Intentional Uses of Memory. *Journal of Memory and Language*, 30, 513–541.

Jäger, G., & Rogers, J. (2012). Formal language theory: refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1956-1970.

Jamieson, R. K., & Mewhort, D. J. K. (2009). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. Quarterly Journal of Experimental Psychology, 62, 550–575.

Jiang, S. (2012). Implicit learning of Chinese tonal regularity. Unpublished dissertation. East China Normal University, China.

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. *Consciousness & Cognition, 21,* 476-486.

Kenny BJ, Gellrich M (2002) Improvisation. In: Parncutt R, McPherson GE, editors. The science and psychology of music performance: Creative strategies for teaching and learning. Oxford: Oxford University Press. pp. 163–180.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing:“nonmusicians” are musical. *Journal of Cognitive Neuroscience*, *12*(3), 520-541.

Koelsch, S., Rohrmeier, M., Torrecuso, R., Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences*. published ahead of print September 3, 2013, doi:10.1073/pnas.1300272110

Kuhn, G., & Dienes, Z. (2008). Learning non-local dependencies. *Cognition, 106*, 184-206.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of non-local musical rules. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*, 1417–1432.

Lai J, Poletiek FH (2010) The impact of adjacent-dependencies and staged- input on the learnability of center-embedded hierarchical structures. *Cognition* 118: 265–273.

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Li, F., Jiang, S., Guo, X., Yang, Z., & Dienes, Z. (2013). The nature of the memory buffer in implicit learning: Learning Chinese tonal symmetries. *Consciousness & Cognition, 22* (3), 920-930.

Lobina, D. J. (2011). “A Running Back” and Forth: A Review of Recursion and Human Language. *Biolinguistics*, *5*(1-2), 151-169.

Loui, P. (2011). Statistical Learning – What Can Music Tell Us? In P. Rebuschat and J. Williams (Ed), *Statistical Learning and Language Acquisition,* 433-462: Mouton de Gruyter.

Loui P, Wessel DL (2008) Effects of Set Size on the Learning and Liking of an Artificial Musical System. Musicae Scientiae 12: 201–218.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco, CA: W. H. Freeman.

Martins, M. D. (2012). Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2055-2064.

Masters., R. S.W., Maxwell., J. P, Eves, F. F. (2009). Marginally perceptible outcome feedback, motor learning and implicit processes. *Consciousness and Cognition, 18*, 639–645.

Miller, G. A. (1967). *The psychology of communication*. Basic Books: New York.

Müller, S. (2010). *Grammatiktheorie*. Tübingen: Stauffenburg Verlag.

Opitz B, Friederici AD (2003) Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. Neuro Image 19: 1730–1737.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*(3), 306-338.

Perruchet, P., & Pacteau, C. (1991). Implicit acquisition of abstract knowledge about artificial grammar: Some methodological and conceptual issues. *Journal of Experimental Psychology: General*, 120(1), 112-116.

Perruchet P, Rey A (2005) Does the mastery of hierarchical structures distinguishes humans from non-human primates? Psychonomic Bulletin & Review 12(2): 307–313. de Vries M, Monaghan P, Knecht S, Zwitserlood P (2008) Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. Cognition 107: 763–774.

Poletiek FH (2011) What in the world makes recursion so easy to learn? A statistical account of the staged input effect on learning a centre embedded hierarchical structure in AGL. Biolinguistics 5: 1–2, 36–42.

Pothos EM (2007) Theories of Artificial Grammar Learning. Psychological Bulletin 133: 227–244.

Reber AS (1989) Implicit learning and tacit knowledge. Journal of Experimental Psychology: General 118: 219–235.

Reber AS (1967) Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior 6: 855–863.

Reber AS (1993) Implicit learning and tacit knowledge: An essay on the cognitive unconscious. New York: Oxford University Press.

Rebuschat P (2008) Implicit learning of natural language syntax. PhD Dissertation. Research Centre for English and Applied Linguistics, University of Cambridge.

Rebuschat, P., & Williams, J. N. (Eds.). (2011). *Statistical learning and language acquisition* (Vol. 1). De Gruyter Mouton USA.

Redington M, Chater N (1996) Transfer in artificial grammar learning: A reevaluation. Journal of Experimental Psychology: General 125: 123–138.

Reed, N., McLeod, P., & Dienes, Z. (2010). Implicit knowledge and motor skill: What people who know how to catch don't know*. Consciousness & Cognition, 19*, 63-76.

Remillard, G. (2008). Implicit learning of second-, third-, and fourth-order adjacent and nonadjacent sequential dependencies. *The Quarterly Journal of Experimental Psychology, 61*, 400-424.

Rodriguez, P., Wiles, J., & Elman, J. (1999). A Recurrent Neural Network that Learns to Count. *Connection Science*, *11*(1), 5–40. doi:10.1080/095400999116340

Rohrmeier, M., & Cross, I. (2013). Artificial grammar learning of melody is constrained by melodic inconsistency. Narmour's principles affect melodic learning. *Plos ONE* 8 (7): e66174

Rohrmeier, M. & Cross, I. (2010). Narmour's principles affect implicit learning of melody. In Demorest et al. (eds.), Proceedings of the 11th International Conference on Music Perception and Cognition (ICMPC 2010).

Rohrmeier, M. & Cross, I. (2009). Tacit tonality: Implicit learning of context-free harmonic structure. In J. Louhivuori et al. (eds.), *Proceedings of the 7th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM 2009)*. pp. 443-452.

Rohrmeier, M., Dienes, Z. , Guo, X. , & Fu, Q. (2014). Implicit learning and recursion. In: Francis Lowenthal & Laurent Lefebvre (Eds), *Language and Recursion*, Springer Verlag.

Rohrmeier, M., Fu, Q., & Dienes, Z. (2012).Implicit learning of recursive, hierarchical grammatical structures. *PLoS ONE 7*(10): e45885. doi:10.1371/journal.pone.0045885

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. Journal of Mathematics and Music, 5 (1), pp. 35-53.

Rohrmeier, M., & Rebuschat, P (2012) Implicit learning and acquisition of music. *Topics in Cognitive Science*, 4 (4), pp. 525-553. DOI: 10.1111/j.1756-8765.2012.01223.

Rohrmeier M, Rebuschat P, Cross I (2011) Incidental and online learning of melodic structure. *Consciousness & Cognition* 20(2): 214-222.

Rohrmeier, M. (2010) Implicit learning of musical structure. Experimental and computational modelling approaches. PhD Dissertation, Centre for Music & Science, Faculty of Music. Cambridge: University of Cambridge, UK.

Rohrmeier, M. (2007) A generative grammar approach to diatonic harmonic structure. In Spyridis, Georgaki, Kouroupetroglou, Anagnostopoulou (Eds.), Proceedings of the 4th Sound and Music Computing Conference, pp. 97-100

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*, 493–515.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*, 172–196.

Saffran JR, Newport EL, Aslin RN (1996) Word segmentation: The role of distributional cues. Journal of Memory and Language 35: 606–621.

Scott, R. B., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 34, 1264-1288.

Servan-Schreiber, D., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 592–608.

Shanks, D. R. (2005). Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202-220). London: Sage.

Steedman, M. (2002). Plans, affordances, and combinatory grammar. *Linguistics and Philosophy*, *25*, 723–753.

Steedman, M. (1996). The Blues and the abstract truth: Music and mental models. In J. Oakhill & A. Garnham (Eds.), *Mental Models in Cognitive Science: Essays in Honour of Phil Johnson-Laird* (pp. 305–318). Psychology Press.

Steedman, M. J. (1984). A Generative Grammar for Jazz Chord Sequences. *Music Perception*, *2*(1), 52–77.

Tomalin, M. (2007). Reconsidering recursion in syntactic theory. *Lingua*, *117*, 1784–1800.

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. *Cognitive Science.* E-pub ahead of print. DOI: 10.1111/j.1551-6709.2012.01235.x

Vokey, J. R., & Brooks, L. R. (1994). Fragmentary knowledge and the processing-specific control of structural sensitivity. *Journal of Experimental Psychology*, 20, 1504–1510.

Westphal-Fitch, G., Huber, L., Gómez, J. C., & Fitch, W. T. (2012). Production and perception rules underlying visual patterns: effects of symmetry and hierarchy. *Philosophical Transactions of the Royal Society B, 367*, 2007-2022.

Williams JN (2005) Learning without awareness. Studies in Second Language Acquisition 27(2): 269–304.

Williams JN (2009) Implicit learning. In: Ritchie WC, Bhatia TK, editors. The New Handbook of Second Language Acquisition. Emerald Group Publishing Limited. pp. 319–353.