To appear as section (7.3) in
Mind in The Mirror of Language, London, Bloomsbury

Syntax and Language Bias: Case Studies in Arithmetic and Music

Nirmalangshu Mukherji
Former Professor of Philosophy
Delhi University

Abstract

There is a considerable literature by now on the generative properties of a variety of non-linguistic systems displayed by humans: arithmetic, music, kinship relations, stone-tool making, artistic products, games like chess, and the like. Given the strong intuition that these are unbounded generative procedures, the question naturally arises whether the basic operation involved in these non-linguistic systems is the familiar linguistic operation Merge.

However, it is not difficult to see that the research in the massively interdisciplinary, and possibly unifying, field has been scuttled by what may be called the *language bias*: either the system under study reflects the basic properties of language or it is not a generative system at all.

I argue that the language bias has no basis even in the paradigmatic cases of arithmetic and music. This is because Merge is not specific to a domain, even if it operates in a specific lexical workspace. So, whether Merge, or its variant, is a universal generative operation is an apt research question.

NOTE: In the book, this discussion is preceded by (a) a long discussion of why Merge is not linguistically specific, and (b) why no notion of computation applies to desert ants and birdsongs.

Syntax and Language Bias

The study of the nonhuman complex systems reveals a problem with the literature that investigates the possibility of assigning generative systems to animals. While examining alleged computational properties of dead reckoning by desert ants, Watumull et al. (2014) made the observation that navigation by ants is not up to the mark for genuine computation since it fails to meet the standards displayed in linguistic recursion. Similarly, Hauser et al. (2014) observed that birdsongs don't qualify as systems of genuine symbolic computation because they do not satisfy the conditions of symbolic computation in human language.

There seems to be a *language bias* in *dis*qualifying non-human systems as computational; a system fails to be computational just because it does not meet the standards of linguistic computation. Beyond mentioning the point, I did not pursue the topic earlier because neither dead reckoning nor birdsongs are examples of generative procedure at all in my view as extensively argued (MML, 7.2). So, whether they meet the standards of linguistic computation is largely irrelevant from that perspective.

However, the topic needs to be addressed now that we are about to explore some cases of human non-linguistic computation systems. The topic is conceptually crucial for this step. The language bias in the literature appears to lead to the following uncomfortable consequence. If some non-linguistic system turns out to be computational by satisfying the conditions of linguistic computation, then they do not really satisfy the crucial condition of displaying computational properties beyond language. If, however, what otherwise looks like computational properties in some non-linguistic system differs from linguistic computation on some measure, then, on that count alone, they are not to be considered as genuine cases of computation;

hence, they fall outside the computational theory of mind.

This consequence is uncomfortable because it becomes virtually impossible to inquire if a certain system of sounds, actions or marks on paper is a computational system unless it is very much like a linguistic system; that is, it is difficult to ask if there are generative systems other than language.[i] As Raffman (1993, 40-1) also points out, we cannot meaningfully ask if there are symbol systems other than language. There are several examples in the literature where such language-bias is directly recommended. I will pick two recent cases.

Citing Chomsky (2000c), Hinzen (2014) points out that something like theta-theory is a property of any language-like system, whereas Case checking theory is specific to human language. Setting aside the Case-theory part, I wish to draw attention to Chomsky's idea that any 'language-like' system, apart from 'human language' itself, contains something like theta-theory. As Hinzen points out, theta-theory is that part of the overall grammatical theory which assigns thematic roles such as agent, goal etc. to arguments (Mukherji 2010, 2.3.1.3).

Thematic roles are arguments that fill the relevant argument-places in the matrix of, say, a verb viewed as a predicate. In the expression, *Bill hit John, Bill* is the agent, *hit* is the role-assigning predicate and *John* is the patient. Basically then, assignment of theta roles requires a predicate-argument structure; specifically, these predicates and their arguments are *linguistically specific* objects such as verbs and nouns. Furthermore, we know that thematic roles relate to the semantic component of languages, where the semantic component is currently viewed as a part of the 'external' CI system. So all this organization has to be in place before we consider any system to be language-like. When such conditions are imposed, it is difficult to see how a non-linguistic system can be 'language-like' while failing to be a language.

Hauser (2016) in fact recommends something like the satisfaction of *Strong Minimalist Thesis* (SMT) to qualify as a 'logic of thought' which includes all four ingredients of linguistic organization: Merge, CI, SM, and the interfaces. For Hauser, the idea of logic of thought is much broader than that of language. As he puts it, the logic of thought consists of 'computations and representations of thought, with linguistic thought being one flavour.' According to Hauser, the logic of thought is 'uniquely human, and underpins not only language, but many other domains as well.' Apparently then, the logic of thought covers all human-specific systems of computation and representation, linguistic or non-linguistic, provided each of them satisfy SMT. Thus, each component of SMT—Merge, CI, SM, interfaces—need to be 'articulated in different domains' to display how the logic of thought works only for humans across the board. Unfortunately, apart from some vague comments on abstract, a-modal concepts and some obscure queries on how Merge might look like neurologically in different domains, Hauser (2016) fails to cite a single human but non-linguistic domain to illustrate the range of logic of thought.[ii]

I am setting aside the vexing issue of what Chomsky (2000c) and Hauser (2016) may have in mind for 'language-like' systems and systems of 'logic of thought' respectively. In my view, the language bias has severely impacted research on human generative devices outside language. I will illustrate this point with quick review of the scene with arithmetic and music.

Arithmetic

Consider Chomsky's remark that arithmetic is an 'offshoot' of human language; at some other places, Chomsky suggests that arithmetic is a 'concomitant of elements that enter into language' (Chomsky et al. 1982). Which elements? Chomsky had made these remarks often. To be fair to Chomsky, he has always made it clear that only the *knowledge* of arithmetic, that is, the study of how the number system may be organized, could have been an offshoot of language; he is not claiming that the number system itself is an offshoot of language. I will argue even that modest claim is unwarranted. To see what he means, let us examine the system of numbers under issue.

By now, it is very well understood that nonhuman animals have at least two forms of number-sense: subitizing and quantitative estimation (Hauser et al. 1996; Gelman and Cordes 2002; Dehaene 2011). Animals have both an exact sense of small sets and a rough estimate of large quantities; these endowments vary for different species and the age of different members within a species. As we saw for desert ants, even they have something like a step counter although researchers warn us that the ants do not literally count. However, if 'counting' amounts to mastering the number system—the system of *ordinals*—then even chimpanzees do not have it.

Although it has been argued that chimpanzees have been trained to recognize individual numerals up to 9, it is unclear if these numerals correspond to the conventional notion of numbers in the cognition of chimpanzees. Thus, Spelke (2011) has argued, most animals have some 'core knowledge' of colour, geometry, numerosity etc. Yet, according to Spelke, these are not the 'abstract concepts' available to humans. It appears to follow that chimpanzees may not possess even the *cardinal* numbers up to 9; the question of possessing *ordinal* numbers does not arise. Spelke holds that language plays a pivotal role in the development of abstract numerical and geometric concepts. So, what notion of language is involved here?

Chomsky is very specific on this point. According to him, 'we might think of the human number faculty as essentially an "abstraction" from human language, preserving the mechanism of discrete infinity and eliminating the other special features of language' (1988, 169). I am setting aside the issue of how abstract mathematical concepts are reached just by adding the mechanism of discrete infinity to the available system of core knowledge. In my view, the problem is exactly on a par with the development of the concept of apple from what we inherited from the chimpanzees (Petitto 2005; Berwick et al. 2013). In this perspective, we cannot take whatever Chimpanzee number sense offers, add the procedure for generating discrete infinity to it, and get ordinal numbers.

In any case, we saw in some detail that the mechanism of discrete infinity for human languages is exactly the mechanism of Merge. Furthermore, Merge is just a blind generator of discrete infinity independently of other special features of language such as the language-specific human lexicon. To emphasize, since Merge itself is not linguistically specific and Merge is the only stuff added to the existing system of numerosity, it is misleading to think of arithmetic as an offshoot of language. We can now see where this misleading idea is coming from: it is coming from the wrong conception that Merge is specific to human language, the FLN hypothesis (Hauser et al. 2002).

The nonlinguistic character of arithmetic is further illustrated by Chomsky's recent suggestion that if we take Merge and we reduce the rest of the generative system to the absolute minimum, that is, a lexicon which contains one element, we get the successor function (Chomsky 2008). According to Chomsky, Merge operates on the minimum of arithmetic, namely a single lexicon 0, and the result is the successor function: 0, (0), ((0)), (((0))), and so on. As we know, if we are able to define the successor function, we can generate the rest of the infinite series of ordinals and some basic arithmetical functions like summation. As an aside, I wish to note that, in suggesting how ordinal numbers may be generated by Merge, we need not assume that Chomsky was making a serious empirical proposal about how children acquire the number system. All he could have meant was that, in order to account for this universal arithmetical ability displayed by children, we need to postulate something like Merge.

In this connection, it is interesting that Ken Hiraiwa (2017) argues that 'linguistic representations of number, grammatical number, and numerals do not incorporate anything like the successor function.' It is not clear what he needs to be convinced that the successor function is 'incorporated'. His argument appears to be that the number system is taken over by the system of numerals in language: he says, 'numeral systems reflect both of the core systems of number and Merge.' Hiraiwa obviously thinks of Merge as a linguistic operation; so, according to him, the discrete infinity reflected in the system of numerals is a linguistic contribution.

However, if Merge is not linguistically specific, then we can simply view Merge as operating on linguistic elements such as numerals to generate the arithmetical continuum of discrete infinity *in* languages in terms of the phonological and orthographic schemes available in them. Otherwise, beginning with the lexicon '0' which is the lexicon representing the first natural number *in* arithmetic, Merge generates the series of ordinals to capture the successor function. This is what Chomsky meant by knowledge of arithmetic and it has little to do with language.

There is interesting evidence that all arithmetic functions remain in place despite severe language impairment (Varley et al. 2005). The study was conducted on three patients who suffered from some difficulty in processing phonological and orthographic properties of number words; that is, the concerned patients had problems in accessing the numeral part of the lexicon of language. The authors report that, otherwise, all 'three patients solved mathematical problems involving recursiveness and structure-dependent operations (for example, in generating solutions to bracket equations)', showing that the computational procedures were intact. This result clearly shows that the computational procedure of arithmetic, namely Merge, remains unaffected despite linguistic impairment.

However, some authors appear to think just the opposite. Thus, Arsenijevic' and Hinzen (2012) argue that the recursion displayed in arithmetic is not Merge because Merge, for them, does not strictly follow the standard x-over-x recursion. As we saw, the recursive generation of ordinal numbers strictly obeys the x-over-x principle, as in Chomsky's scheme. In contrast, we saw in Chapter Six that generative structures of language typically require incorporation of different syntactic categories in order to complete the phase: $[_{CP}$ Allegedly, $[_{TP}$ John will $[_{vP}$ deny $[_{DP}$ the very possibility $[_{CP}$ that $[_{TP}$ . . .]]]]]] (see 6.4.2).

Arsenijevic' and Hinzen's argument is a clear demonstration of language-bias in their understanding of Merge. Merge in the language case is taken to be the real Merge such that if any computational operation differs from the language case, it is no longer Merge. Once we delink the idea of Merge from its domain-specific operations in language, the basic operation could be viewed as more abstract than any of its specific local applications. Of course, we need to show carefully how the products of Merge vary in structure in terms of the demands of the lexicon in respective domains, but that is a separate issue. Here my only concern is to point out the language bias in the decision to characterize Merge in terms of the specificity of its operations in the domain of language, rather than in arithmetic.

In this connection, it is interesting to note how nonlinguistic arithmetic really is. For example, there is nothing like thematic roles in arithmetic which Chomsky thinks is a property of any language-like system. Furthermore, even if arguably the discourse of arithmetic satisfies Hauser's criteria of Logic of Thought in terms of Merge and the two interfaces, the sound-meaning corelation for arithmetic is likely to be direct. This is because the lexicon of arithmetic does not contain the required features for driving Internal Merge, as we saw with metrical stresses (Berwick 2011). Arithmetic then is best viewed as a kindred system of language rather than as a language-like system in Chomsky's narrow sense; it is certainly not an offshoot of language.

The kindred system of metrical stress suggests another basic difference from the language system. Despite all its differences, arithmetic may still be viewed as akin to language because the discourse of arithmetic—for example, '5 > 3'—involves propositions which admit of the standard Tarski-type satisfaction conditions, that is, truth-conditions. In contrast, the system of metrical stress with its 'impoverished lexicon' of x and / (see Chapter Six), does not seem to have such affiliations with the language system. It is meaningless to think of a sequence of metrical stresses as a proposition with some truth value attached. In that sense, the system of metrical stress is further away from language than arithmetic. All we can say is that human languages, arithmetic and the system of metrical stress belong to a kindred class exactly because they are all Merge-based systems.

Music

The domain of music, another discretely infinite system, seems to be a more interesting candidate for the suggested class of kindred systems precisely because it is phenomenally almost totally distinct from the domain of language. Locating the same computational system in two phenomenally distinct domains promises significant theoretical unification towards the concept of mind. In what follows in this section on music, I will be mostly focusing on the phenomenal differences between language and music to insist that the only thing they have in common is the computational system.

In fact, as we will see, the organization of music is indeed like birdsongs in appearance with the exact difference that human music is a genuinely generative system of I-computation.

Although I have written about these differences between language and music earlier at length (Mukherji 2000; Mukherji 2010), I wish to emphasize some of the crucial points because, unlike the focus on differences here, earlier I was also anxious to stress the stark similarities between the two systems to suggest something like a musilanguage hypothesis, even a strong musilanguage hypothesis. At that point, it seemed to me that showing Merge-like computational operations to explain the generativity of music was the only convincing way of demonstrating the domain-independence of Merge within a class of language-like systems. Now that we have Principle C in hand and a strong sense of separation between the lexicon and the computational system, we no longer have the preceding anxiety about the domain-independence of Merge (Mukherji 2021): Merge *is* domain-independent whether it applies to music or not. Therefore, there is no need for a specific musilanguage thesis. The application of Merge in the domain of music simply expands the scope of Merge.

The generative character of human music is pretty obvious and was recognized for ages. Every musical system consists of a small set of notes and syllables as units that enter into computation. These units are compiled over and over again to generate progressively complex objects such as chords, scales, modes, phrases, passages, movements, and so on; as we will see, some of these are viewed as forming the lexicon of music. The generation of non-lexical complex objects is unbounded, hierarchical, and countable (Brown 2000, 273; Merker 2002; Fitch 2006). Even then authors continue to think that 'there are no unambiguous demonstrations of recursion in other human cognitive domains, with the only clear exceptions (mathematical formulas, computer programming) being clearly dependent upon language' (Fitch et al. 2005). I return to the remark below.

7.3.1 Merge in music

As to the recursive character of progression in music, Pesetsky (2007) suggests that musical progression can be displayed as structures with binary branching and headed phrases. They are thus characterizable as products of Merge. To establish the point, Pesetsky redraws the diagram for the progression of the opening part of Mozart's piano sonata K. 331 (Figure 7.2).

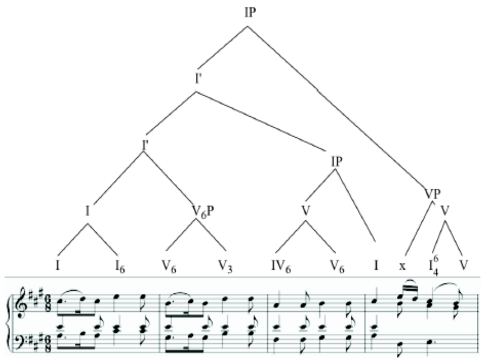

Figure 7.2 Merge in music. Opening part of Mozart Piano Sonata K. 331.
Reproduced from Mukherji (2010), Originally in Pesetsky (2007).

Many more diagrams for a variety of tonal music were offered in Katz and Pesetsky (2011). Similar structures can be generated for Indian classical music (Mukherji 2010) in the set notation. Skipping details of Pesetsky's diagram, the basic idea is to treat 'I' and 'V' as heads of phrases, where 'I' is the tonic chord and 'V' the dominant chord.[iii]

Before we proceed, I wish to point out a general problem with musical notation as with Berwick's notation for metrical stress in Berwick (2011). As we saw there, the symbols x and / cannot be the symbols on which Merge operates; these are Berwick's inventions for making a theoretical point. Similarly, for Pesetsky's notational scheme. Even if 'I' and 'V' are taken to be theoretical notation like NP and VP as in the language case, the stave notation for notes are not the symbols on which Merge operates. In the language case, in contrast, the syntactic structure '{*saw* {*the man*}' contains words of English, which are not artificial notations.[iv] These words are symbols as they are the articulated forms of mental representations; thus, Merge can operate on these symbols. The artificial notation in music at best suggests that certain mental representations are involved, but the notation does not specify what these mental representations are. These mental representations then are the symbols on which Merge operates. Near the end of this section, I engage in some speculation on what these mental representations are likely to be.

Returning to other aspects of Pesetsky's diagram, when a sequence is generated by Merge, as we saw with metrical stress, it is natural to assign some selections even from the 'impoverished' lexicon as heads with projections to show the growth of the structure. As we saw, the online construction of head-complement structure is just a natural consequence of Merge; it has nothing to do specifically with the language system. My claim is that, beyond this structural convergence due to Merge, there is nothing linguistic about music. Which input items are selected as heads and how the projections are to be interpreted depends entirely on the concerned domain.[v]

Hence, I am not at all convinced about what Katz and Pesetsky call the 'Identity' between language and music. Again, the system of metrical stress can be used as a point of departure. For the system of metrical stress, Berwick remarked that it was an example of syntax 'with no lexical features and no associated semantics; without features, there is no possibility of internal merge and the movement that we see in ordinary syntax'; he calls it 'pure syntax' (464). It is unclear why the system of metrical stress should be called *syntax* at all when it has no other language-like features, especially with no associated semantics. If, however, by 'syntax' Berwick means just the presence of the generative operation Merge, then of course there is no substantive problem. Similarly, for music.

In my view, the very fact that Pesetsky was able to describe musical progression with Merge, then other things being equal, it refutes the language-bias of Fitch et al. (2005) that 'there are no unambiguous demonstrations of recursion in other human cognitive domains.' More significantly, a demonstration of convergence between language and music in this very specific generative respect finally answered the long-standing problem of explaining unbounded, hierarchical, countable generativity in music. To my knowledge, no other form of work in music theory has a satisfactory answer to this problem; for example, the otherwise classic work of Lerdahl and Jackendoff (1983) suffers from this defect (Mukherji 2010).[vi] On that count, if we wish to call Pesetsky's Merge-based account of musical progression 'musical syntax', there is no problem. We can still uphold the view that language and music share nothing else.

However, to motivate the 'identity thesis' between language and music, Katz and Pesetsky claim more. According to them: 'All formal differences between language and music are a consequence of differences in their fundamental building blocks. In all other respects, language and music are identical.' As the authors clarify, the 'building blocks' of language consists of lexical items which are arbitrary sound-meaning pairings, while music consists of pitch-classes (roughly, notes) and combinations of pitch-classes such as chords, scales, ragas, motifs etc. Later in the section, I discuss what these things mean in mental terms so that Merge can operate on them; the discussion will thereby indicate how musical thought looks like.

As we know, the arbitrary pairing between sound and meaning in the language case ultimately receives interpretation of complex sequences at two independent interfaces of SM and CI. So, we have two representations, say, PHON and SEM and the pairing between them, <PHON, SEM>. The syntax of human language is such that it generates the pair. In the musical case, however, there is no such arbitrariness and independence between pitch-classes and, say, chords since pitch-classes (notes) constitute the chords. In that way, what constitutes the lexicon of music is more like units of phonology and morphology which constitute PHON. This raises the question if music has semantics at all in any interesting sense.[vii] Since Katz and Pesetsky do not discuss the question, I will discuss it later. Setting semantics aside then for now, what remains in music syntax, beyond Merge, that enables the authors to claim the formal identity thesis?

As far as I can see, the only other aspect of syntax pointed out by Katz and Pesetsky concerns movement or displacement of syntactic objects by Internal Merge. As we saw, Internal Merge is just Merge and it comes for free.[viii] Yet, the availability of some feature does not mean that it is actually availed of. Recall that movement in the language case is triggered by some morphological feature of the lexical head; also, movement consists of copies. So, there are morphological properties that force placement of copies at vacant places such that some syntactic requirement, such as agreement, may be met. The operations of Internal Merge produce the phenomenon of displacement: a syntactic object is heard at one place while it is interpreted somewhere else. Hence, sound and meaning are mostly indirectly related in the language case. It is doubtful whether these aspects of movement are simulated in the music case. Berwick (2011) denied movement in metrical stress precisely because the heads in metrical stress were featureless. However, let us suppose Berwick reached that conclusion due to his language-bias: nothing moves unless it moves in the way syntactic objects move in language.

Katz and Pesetsky (2011) argue for a different form of movement in music. They suggest that cadences in compositions of music may be viewed as products of Internal Merge.[ix] Western classical tonal music typically imposes an adjacency condition: some chords of certain types must be adjacent to meet the phrase closure requirement; the adjacent chords form a cadence (Mukherji 2013). According to Katz and Pesetsky, syntactic objects in music sometimes move to meet the adjacency constraint on phrases and, in so moving, they form cadences. Several problems with this explanation of movement in music has been pointed out. For example, cadences for phrasal closure are typically found in Western classical tonal tradition, they are not found in rock music; hence such cadences are not universal features of music, probably they are imposed by idioms. A more formal objection is that it is not evident that in the suggested cases some syntactic movement is actually taking place (Mukherji 2013).

Be that as it may, since there are no cadences in human languages, the suggested case of Internal Merge in music, even if it is valid, is not an evidence in support of the identity thesis. It is clear where the Katz-Pesetsky identity thesis is coming from: it is coming from a language-bias that Merge is a linguistic operation and hence any evidence of Internal Merge is further evidence that language and music are identical. The assumption that Merge is dedicated to language is essentially false, as we saw.

The point that language and music are formally entirely different can be advanced further if we consider the notion of musical interpretation vis-à-vis linguistic interpretation. Although, Katz and Pesetsky do sketch how the generative structure of music looks like, they do not say anything about how the interfaces, where these structures are interpreted, look like. Assuming that musical organization also contains interfaces with ‘external’ systems in some sense, they cannot be the external systems of language because interpretation of music centrally involves the tonal organization of music, not the phonetic and semantic representations of language.[x]

7.3.2 Significance in music

In fact, any form of identity thesis between language and music is suspect because of the predominance of the sound component in music. As we saw, if we wish to pursue the analogy of language-design, we can at best say, if at all, that music consists only of a PHON-like component. In the absence of an overt SEM component, music is starkly different from language. Indeed, the recent proposal that the sound-part of language is ancillary (Berwick and Chomsky 2016) could mean that language is essentially geared to meet the conditions at the CI interface. There is no such thing as CI interface for music; musical objects do not express any 'external' concepts or refer to any item in the world. In that sense, musical sequences do not *say* anything which could be expressed in words.

Sometimes this phenomenon is misleadingly characterized as an 'ineffability problem' (Raffman 1993). Ludwig Wittgenstein clarified why it is misleading to view the alleged property of music as a problem: 'it is a strange illusion that possesses us when we say "This tune says *something*", and it is as though I have to find *what* it says' (Wittgenstein 1931/1958, 178). As Wittgenstein characteristically suggested, the solution to the alleged ineffability problem is to whistle the tune. Since I have written extensively on this topic (Mukherji 2010), I will not pursue it any further here. Nevertheless, I will mention one specific issue briefly to draw the analogy with birdsongs.

The language bias, pointed out by Wittgenstein to address the so-called ineffability problem, explains why authors devote so much effort to locate some 'external' system in the music case as well. Since it is well understood by now (Fitch 2006; Boghossian

2007) that we cannot associate a musical sound with either a concept or something in the world, music is best understood as expressing emotions. Authors such as Jackendoff and Lerdahl (2006) often come up with elaborate list of these emotions: gentle, forceful, awkward, abrupt, static, earnest, reverent, tender, ecstatic, sentimental, longing, striving, resolute, depressive, playful, witty, ironic, tense, unsettled, heroic, etc. (see Mukherji 2010 for more).

The basic problem, among many, is that even if we are able to ascribe specific emotions to given pieces of music, emotions can only be global properties of music. Typically, emotions are assigned to an entire piece of music; for example, according to Roger Scruton, the last movement of the *Jupiter Symphony* is 'morose and life-negating' (cited in Raffman 1993, 42). But music is a complex organization of notes which we hear on a note-by-note basis, forming larger and larger groups as we proceed. It is hard to see how the global (emotional) property of the piece is computationally reached from its smaller parts; emotions just don't compute in the desired sense (Mukherji 2010). But this is not to deny that expression of emotions is a significant aspect of music.

The point of interest here is that expression of emotions as global properties of large chunks of music is exactly how birdsongs are associated with some global significance for the species, as we saw. In fact, songbirds sometimes are said to display songs out of pleasure. If pleasure falls under emotions, then birdsongs are also associated with emotions as global factors. Now we can see why it is misleading to complain, both for birdsongs and human music, that 'when song syllables are combined to create longer structures, new creations have no impact on the function or "meaning" of the song' (Hauser et al. 2014).

The point is, these global associations are not to be viewed as features of computation at all; in that sense, their task *is* to serve a global function. The intimate association between music and emotions holds without tracing the association to the internal structure of music. As songs—that is music with lyrics—attest, outputs of both language and music certainly access emotions. It is a common experience that most complex human activities are globally associated with emotions: watching soccer or cooking 'Mama's pasta', for example. To my knowledge, it is not very clear why such associations take place at all.

For the music case, one speculation is that music accesses emotions more directly and definitively because music is fundamentally an 'internal' system; music has nothing else to access (Mukherji 2010). For another explanation, Dan Sperber (1994) suggests that the human music 'module' could have been a cultural adaptation of 'acoustic properties of early human vocal communications.' According to Sperber, 'sounds that the module analyzes thereby causing pleasure to the organism … are not often found in nature (with the obvious exception of bird songs).' The mention of birdsongs in this rather exclusive context suggests that we can view human music as an adaptation of some pre-existing vocal system that we shared with songbirds.

In any case, the critical difference between the two is that human music is generative in character, birdsongs are just patterns of complex sound. Under the simplest assumption that some version of Merge instantiates Principle C, it is not surprising that musical form is the result of successive applications of Merge, as Katz and Pesetsky showed. But Merge is a computational device that needs some lexical workspace constituting of symbols. In the absence of conceptual, referential and emotional properties in musical meaning, what could be the symbols in the music case?[xi]

No doubt, the most prominent part of interpretation in music concerns internal relations between musical elements themselves without any computational connection with external factors such as concepts and emotions. As Diana Raffman (1993, 59) explains the phenomenon, musicians are typically concerned with *structural* issues such as whether a given phrase ends at E-natural. Three possibilities arise: the phrase ends before the E-natural, the phrase ends at the E-natural, and the phrase extends beyond the E-natural. As anyone familiar with music knows, these structural variations make substantial differences in the interpretation of music. Depending on the group of notes at issue, and the location of the group in a passage, some of the structural decisions may even lead to bad music. This is because these decisions often make a difference as to how a given sequence of notes is to be resolved. Any moderately experienced listener of music can tell the differences phenomenologically, though its explicit explanation requires technical knowledge of music (such as modulation to the dominant).

In my view, this aspect of musical interpretation is on a par with much of grammatical interpretation in language since, at the level of grammar, linguistic interpretations are restricted to structural issues such as scope distinction, distribution of arguments, agreement between syntactic objects, and the like (Mukherji 2010). Arguably, no mention is made at this level of the conceptual, referential and other 'external' features of lexical items.[xii] But it would be facile to draw any deep connection between language and music on this count. It could be just a reflex of Merge-like operations that lexical items, and groups of them, form various structural relations at different hierarchies. Which of these structural relations are salient and which violate conditions on interpretation depends on the concerned domain. For example, in the language case, grammatical computation has to satisfy conditions at the semantic interface. It does not follow that music also needs to have something like semantic representation in its overall grammatical organisation, as suggested in Mukherji (2010).

In any case, the query about individual symbols in music persists because, after all, most lexical items of language receive progressively wider interpretation in the vast CI system, as the so-called semantic features associated with lexical items display. In the absence of a CI system in the music case, the notion of a musical symbol thus poses a bit of a problem if we suffer from language bias in our concept of a symbol. In the absence of a language-like CI interface for music, some authors indeed suggest that music is not a symbol-system at all. As noted, the trouble is that Merge can operate only in a symbolic domain.

Setting aside the history of language bias in this area, one way to approach the problem is to ask: what are the constituents of musical thought? What is the composer 'looking up' when he is writing a piece of music? What does a competent user of music hold in the mind when there is no sound in a context? No doubt, a composer occasionally hits some keys on the nearest piano to elicit the sound he wants. But suppose there is no sound of the piano and no associated humming, as we are told happened with Beethoven in his later years. Needless to say, Beethoven is mentioned just to indicate the soundless context of explanation; explanation of how the mind of a musical genius works is an entirely different issue, if feasible.

To explain the content of musical thought in such noise-free contexts, it is plausible to suppose that the musician, as well as the experienced music listener, have some abstract images of the basic sounds stored. These images are 'abstract' in the sense that they form (a) pitch-classes in the mind, say, C in one notational scheme, and (b) form collection of pitch-classes, say, chords that hold across transpositions. In both cases, the image remains invariant while the stimulus, the actual sound, changes, say, as octaves change.

The abstract character of these images suggest that they are made up not only of acoustic properties, but involve some cognitive categories as well (Krumhansl et al. 2000, 14). My guess is that even lay listeners have some vague images of musical sound as they track intonation and sameness across a vast range of common music. They may not have stable images for individual notes, and fail to appreciate idiom-specific structural nuances especially in classical music, but they can recognize whether the intended note has been reached or the same group of notes have been transposed to a different octave, perhaps even in the transposition to a different key.

Even though these images are abstract in the sense indicated, there is no mystery about them since they are fundamentally guided—but not wholly determined—by physical principles of sound such as certain parts of the Fourier series. Roughly, when the auditory receptor system is able to identify the units in the audible range of a Fourier series, and is able to associate group of units in terms of consonance and dissonance, the store of musical sounds is formed. I am assuming roughly the same mechanism for the production of sound when other physical conditions are fulfilled. Obviously, what is stored in the mind are not the sounds themselves, but a set of abstract images of sound as suggested; much of the rest of music is arts, a cultural adaptation as Sperber suggests.

These quasi-stimulus-dependent images then are the basic representations of musical sound in the mind: ‘stimulus-dependent’ because these are images formed from stimulus of sound, ‘quasi’ because these are abstract but not as abstract as concepts. Since these images are naturally formed as suggested, it is likely that some prehuman hominids met the receptor conditions to form such images. The finite collection of basic musical images then could have been available, on a par with the rudimentary set of concepts for language, before the 'long train of thought' started for both music and language, as Charles Darwin (1857) suggested. My contention is that these mental representations of musical sound, in the form of quasi-stimulus-dependent images, are the symbols on which Merge in music operates; these imagistic symbols stand for musical sounds, pitch-classes, in thought.

In some cultures, especially those with written notation, these symbols are associated with phonetic or orthographic forms such as A, B, C, D, E, F, G or do-re-mi-fa-so-la-ti or sa-re-ga-ma-pa-dha-ni; sometimes they are indicated in more technical notations such as the stave notation, Pesetsky's chord notation, etc. But these are not concept-bearing names, these are markers of (stored) images of sound, on a par with Berwick's invention of x and / to mark metrical stress (Berwick 2011). These notational devices play a role in the analysis of music, but they are not needed in the music itself. They do not constitute musical thought.

To say this is not to deny that the vast range of musical concepts such as tonic, dominant, chord, melody, rondo, allegro etc. can enter into discussion of music in the language of music theory. Human languages are used to talk about almost anything, including human musical experiences. For those who are versed in the vocabulary of music theory, these concepts indicate what a sequence of notes 'says'. I am not denying that they also enter into (advanced) musical thinking as an assistance to, say, composition or recall of music, for those who are competent with these notational schemes.[xiii] But they do not enter the composition itself. Classical musicians in India sometimes indeed use the notion of 'saying' by showing how a musical piece sounds like, that is what the piece 'says'. As Wittgenstein pointed out, all you need to do to display the symbolic form of musical thought is whistle.

Symbols of music thus incorporate imagistic information by accessing the relevant perceptual system. But the non-conceptual, imagistic nature of information in music does not prevent music to attain its generative power. The generative power of music comes from the operations of Merge; for Merge, it does not matter if it operates on images or concepts as long as some information is packaged in a symbolic form. Except for Merge, the musical system is fundamentally different from the language system.

References

Arsenijevic', B. and W. Hinzen. 2012. On the absence of X-within-X recursion in human grammar. *Linguistic Inquiry*, 43.3, 423–440.

Asano, R. and C. Boeckx. 2015. Syntax in language and music: What is the right level of comparison? *Front. Psychol.* 6, 942. doi:10.3389/fpsyg.2015.00942

Berwick, R. 2011. All you need is Merge. In A. M. Di Sciullo (Ed.) *Biolinguistic Investigations*. Oxford: Oxford University Press, 461-491.

Berwick, R., A. Friedrici, N. Chomsky, and J. Bolhuis. 2013. Evolution, Brain and the Nature of Language. *Trends in Cognitive Science*, 17.2.

Berwick, R., and N. Chomsky. 2016. *Why Only Us: Language and Evolution*. Cambridge, Ma.: MIT Press.

Boghossian, P. 2007. Explaining musical experience. In K. Stock (Ed.) *Philosophers on Music*: *Experience, Meaning, and Work*. Oxford: Oxford University Press.

Brown, S. 2000. The 'musilanguage' model of music evolution. In N. Wallin, B. Merker and S. Brown (Eds.) *The Origins of Music*. Cambridge, Ma.: MIT Press, 271-300.

Chomsky, N. 1988. *Language and the Problem of Knowledge: The Managua Lectures*.

Cambridge: MIT Press.

Chomsky, N. 2000. Minimalist inquiries. In R. Martin, D. Michaels, J. Uriagereka (Eds.), *Step by Step*. Cambridge, Ma.: MIT Press, 89–155.

Chomsky, N. 2008. On phases. In R. Freidin, C. Otero and M. Zubizarreta (Eds.). *Foundational Issues in Linguistic Theory. Essays in Honor of Jean-Roger Vergnaud*, 133–166. Cambridge, Ma.: MIT Press.

Chomsky, N. 2020. *The UCLA Lectures*. https://lingbuzz.net/lingbuzz/005485.

Chomsky, N., R. Huybregts and H. Riemsdijk. 1982. *The Generative Enterprise*. Dordrecht: Foris.

Dehaene, S. 2011. *The Number sense: How the Mind Creates Mathematics*. Oxford: Oxford University Press.

Fitch, W. 2006. The biology and evolution of music: A comparative perspective. *Cognition*, 100, 173–215.

Fitch, W., M. Hauser and N. Chomsky. 2005. The evolution of the language faculty: clarifications and implications. *Cognition*, 97.2, 179-210.

Gelman, R. and S. Cordes. 2001. Counting in animals and humans. In E. Dupoux (Ed.) *Essays in Honor of Jacques Mehler*. Cambridge, Ma.: MIT Press.

Hauser, M. 2016. Challenges to the what, when, and why?. *Biolinguistics*, 10, 001–005.

Hauser, M., P. MacNeilage, M. Ware. 1996. Numerical representation in primates. *PNAS*, 93.4, 1514-1517.

Hauser, M., N. Chomsky, W. Fitch 2002. The faculty of language: what is it, who has it, and how did it evolve?. *Science*, 298, 1569-79.

Hauser, M., C. Yang, R. Berwick, I. Tattersall, M. Ryan, J. Watumull, N. Chomsky and R. Lewontin. 2014. The mystery of language evolution. *Front. Psychol.*, 07 May, doi: 10.3389/fpsyg.2014.00401.

Hauser, M. and J. Watumull. 2017. The universal generative faculty: the source of our expressive power in language, mathematics, morality, and music. *Journal of Neurolinguistics*, 43, 78-94.

Hinzen, W. and M. Sheehan. 2013. *The Philosophy of Universal Grammar*. Oxford: Oxford University Press.

Hiraiwa, K. 2017. The faculty of language integrates the two core system of numbers. *Front. Psych*. 8. doi: 10.3389/fpsyg.2017.00351.

Jackendoff, R. and F. Lerdahl. 2006. The capacity for music: what is it, and what's special about it? *Cognition*, 100.1, 33-72.

Katz, J. and D. Pesetsky. 2011. The identity thesis for language and music. https://ling.auf.net/lingbuzz/000959

Krumhansl, C., P. Toivanen, T. Eerola, P. Toiviainen, T. Järvinen and J. Louhivuori. 2000. Cross-cultural music cognition: cognitive methodology applied to North Sami yoiks. *Cognition*, 76, 13-58.

Lerdahl, F. 2013. Musical syntax and its relation to linguistic syntax. In M. Arbib

(Ed.), *Language, Music, and the Brain*: *A Mysterious Relationship*. Strüngmann Forum Reports, 10. Cambridge, Ma.: MIT Press.

Lerdahl, F. 2015. Concepts and representations of musical hierarchies. *Music Perception*, 33, 83-95.

Lerdahl, F. and R. Jackendoff 1983. *A Generative Theory of Tonal Music*. Cambridge, Ma.: MIT Press.

Merker, B. 2002. Music: The missing Humboldt system. *Musicae Scientiae*, 6, 3-21.

Mukherji, N. 2000. *The Cartesian Mind*: *Reflections on Language and Music*. Shimla: Indian Institute of Advanced Study.

Mukherji, N. 2009. Dawn of music theory in India. In R. Singh (Ed.) *Trends in Linguistics, Annual Review of South Asian Languages and Linguistics*. The Hague: Mouton de Gruyter.

Mukherji, N. 2010. *The Primacy of Grammar*. Cambridge: The MIT Press.

Mukherji, N. 2021. Preface to *Mind in the Mirror of Language*. https://ling.auf.net/lingbuzz/005850

Mukherji, S. 2013. *Generative Musical Grammar: A Minimalist Approach*. Ph. D thesis. Princeton University.

Patel, A. 2008. *Music, Language, and the Brain.* New York: Oxford University Press.

Patel, A. 2019. Evolutionary music cognition: Cross-species studies. In P. Rentfrow and D. Levitin (Eds.) *Foundations in Music Psychology: Theory and Research*, 459-501. Cambridge, MA: MIT Press.

Peretz, I. 2001. The biological foundations of music. In E. Dupoux (Ed.) *Language, Brain, and Cognitive Development*: *Essays in Honour of Jacques Mehler*. Cambridge: MIT Press, 435-45.

Pesetsky, D. 2007. Music syntax is language syntax. Ms. web.mit.edu/linguistics/people/ faculty/pesetsky/Pesetsky_Cambridge_music_handout.pdf (see also Katz and Pesetsky).

Petitto, L-A. 2005. How the brain begets language. In J. McGilvray (Ed.), *The*

Cambridge Companion to Chomsky, 84–101. Cambridge: Cambridge University Press.

Raffman, D. 1993. *Language, Music and Mind*. Cambridge, Ma.: The MIT Press.

Spelke, E. 2011. Natural number and natural geometry. In S. Dehaene and E. Brannon. *Space, Time and Number in the Brain*. Academic Press, Chapter 18, 287-317.

Sperber, D. 1994. The modularity of thought and the epidemiology of representations. In L. Hirschfeld and S. Gelman (Eds.). *Mapping the mind: Domain specificity in cognition and culture*. Cambridge: Cambridge University Press.

Varley, R., N. Klessinger, C. Romanowski, and M. Siegal. 2005. Agrammatic but numerate. *PNAS*. www.pnas.org_cgi_doi_10.1073_pnas.0407470102.

Vijaykrishnan, K. 2007. *The Grammar of Carnatic Music*. Berlin: Mouton de Gruyter.

Watumull, J., M. Hauser, I. Roberts and N. Hornstein. 2014. On recursion. *Frontiers in Psychology,* 08 January. doi: 10.3389/fpsyg.2013.01017.

Wittgenstein, L. 1931/1958. *The Blue and Brown Books*. Oxford: Blackwell.

Notes

[i] In my view, the language bias has seeped into the literature due to the pioneering work by Alan Turing and the immensely influential work on language by Noam Chomsky. As discussed in Chapter Three, the very idea of computationalism was initially founded on such language-based ideas.

[ii] Similar considerations appear to lead Hauser and Watumull (2017) to suggest that there is a 'universal generative faculty' (UGF) that governs the domains of language, mathematics, morality, and music. However, apart from the familiar general suggestion that the universal generative faculty is something like a Universal Turing Machine and that Merge 'may also play a role in moral judgment' etc., the authors fail to offer any principled explanation of why the UGF covers exactly the four domains mentioned. There are a variety of other problems with the proposal.

[iii] By drawing attention to the structural convergence, I am not claiming that Pesetsky's selection of musical heads is correct or that the piece of music has the suggested form; that is for the music-theorists to decide (Mukherji 2013; Lerdahl 2013; Lerdahl 2015).

[iv] In my view, the absence of overt word-like units in music is another reason for language bias in this area.

[v] Needless to say, for the purposes of this project, I am only concerned about the application of Merge in musical contexts; hence, I am not concerned about the study of music cognition in general. So, I will not comment on the voluminous and complex literature on the overall organisation of music from the multidisciplinary dimensions of acoustics, evolution, neural architecture and cognition (Peretz 2001; Fitch 2006; Patel 2008; Asano and Boeckx 2016; Patel 2019 etc.)

[vi] Although Lerdahl and Jackendoff (1983) assure that 'the process of subordination (or domination) can continue indefinitely' (p. 13) and describe how 'recursive elaboration' may look like in music (p. 214), there is no principled explanation of why these things happen.

[vii] The absence of language-like semantics in music leads some authors such as Vijaykrishnan (2007) to conclude that music has only 'phonetics'. For Carnatic music, he declares that it has neither syntax nor semantics. I have discussed Vijaykrishnan's optimality theory of Carnatic music in Mukherji (2009).

[viii] In more recent work, Chomsky (2020) suggests that IM is in fact more fundamental since it is less costly as it operates on a restricted workspace. I am setting these suggestions aside.

[ix] Pesetsky (2007) argued that the transfer of a syntactic phrase to the rhythmic phrase required Internal Merge. See Mukherji (2010) for a criticism of this argument.

[x] It follows that, if anything, musical structures require something like interpretive systems internal to the faculty of music. In Mukherji 2010, Chapter Seven, these systems were called 'interface conditions driven by the faculty of music', FMI.

[xi] Interestingly, as discussed in detail in Mukherji (2010), a similar question arises for the contribution of lexical elements in grammatical computation for language as well.

[xii] 'Arguably' because some authors do hold that grammatical computation proceeds with an eye on referential properties to be established within a wider notion of grammar (Hinzen and Sheehan 2013). I am setting this suggestion aside because it is unclear to me if the authors work with the same conception of grammatical organization of language.

[xiii] Recall the fascinating scene near the end of Foreman's film *Amadeus* where Mozart on his death-bed dictates his last music to Salieri. Mozart was portrayed as

---

uttering things like *C Major*, *measure five*, etc. What was Mozart dictating these things from?