Seven Theses on the Biology of Music and Language

Bjorn Merker
Seven Theses on the Biology of Music and Language

Bjorn Merker

Introduction

The nature and origin of music, its relation to language, and to our biology more generally have been active topics of inquiry and debate at least since Darwin and Spencer pondered them in the nineteenth century (Darwin 1871; Spencer 1911). They are still subject to both debate and inquiry (Wallin et al. 2000; Honing et al. 2015). As a contribution to this Signata special issue on “Music and Meaning” I present seven theses on the biology of music and language, with annotation for each thesis.

The theses themselves are given in compressed form, and then expanded upon in an annotation following each thesis. These annotations do not present comprehensive treatments of any given thesis, but rather are intended to explicate each thesis sufficiently to make its compressed form comprehensible, and to provide the reader with key references to relevant treatments of its topic in the literature.

Taken together the theses represent a summary of my conceptual commitments regarding topics that I perceive to have received insufficient attention or even to have suffered outright neglect in our attempts to understand the nature and biology of music and language. Rather than embodying a theory of the biology of music and language (for which see Merker 2012) they are intended as a primer of facts and concepts that are all too easily overlooked in fashioning such a theory.
**Thesis 1: Generativity**

The most fundamental property shared by human language and music is their membership in the small class of generative systems based on the so-called “particulate principle”, by which unlimited pattern diversity is generated through combinations among a finite (typically small) set of non-blending elements (von Humboldt 1836; Abler 1989; Merker 2002). This, rather than recursion, is the key to the unbounded generativity of music and language.

Annotation

In a too seldom noted 1989 paper, William Abler combined Wilhelm von Humboldt’s idea regarding the combinatorics of human language with the non-blending principle of genetic recombination from Roland Fisher to define the “particulate principle of self-diversifying systems” (Abler 1989). Abler highlighted three principal instantiations of such “Humboldt systems”: chemistry and genetics employing physical particles (atoms and nucleotides, respectively), and human language employing conventional particles (phonemes) for combinatorial purposes. Abler mentioned music in passing in this connection but did not elaborate. Music does, however, fully qualify as a second human Humboldt system on account of its structural and generative characteristics, as detailed in my paper “Music: The missing Humboldt system” (Merker 2002).

![Cartoon illustration of the “particulate principle of self-diversifying systems”, following Abler (1989).](image)

- (a) A “blending system” in which combining ingredients average. Here exemplified by a drop of ink in water: the combining elements do not generate a qualitatively new entity. Other examples are most mixtures of liquids as well as gases, as in weather systems, and patterns of heat conduction.
- (b) A “particulate system”, in which the combining elements generate a qualitatively new entity by retaining their individuality on combining.
- (c) A miniscule sample of the infinite generativity of a combinatorics of as few as one or two discrete non-averaging elements.

In such systems the combining elements must be non-blending in the sense of not producing an average when combined (Fisher 1930), i.e. they must retain their individuality on combining (Figure 1a, b). When that is the case, each such
combination “creates something which is not present per se in any of the associated constituents” (von Humboldt 1836, p. 67), making infinite pattern variety possible (Figure 1c). This is the badly named “digital infinity” of Chomskyan linguistics, where the “digital” term stands for the discrete, non-blending aspect of particulate combinatorics. When Chomsky stated that “Language is, at its core, a system that is both digital and infinite. To my knowledge, there is no other biological system with these properties” (Chomsky 1991, p. 50), he was only displaying the limits of his knowledge, because a total of four major such systems are in fact flourishing in and all around us. Two of these are independent of humans (chemistry and genetics), while two of them lie at the heart of human culture (music and language), as itemized in Table 1.

<table>
<thead>
<tr>
<th>System</th>
<th>CHEMISTRY</th>
<th>GENETICS</th>
<th>MUSIC</th>
<th>LANGUAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>product</td>
<td>All molecules</td>
<td>All life forms</td>
<td>All “melodies”</td>
<td>All sentences</td>
</tr>
<tr>
<td>Combining “particles”</td>
<td>atoms</td>
<td>nucleotides</td>
<td>Notes</td>
<td>phonemes</td>
</tr>
<tr>
<td>Particle type</td>
<td>physical</td>
<td></td>
<td>conventional</td>
<td></td>
</tr>
<tr>
<td>domain</td>
<td>nature</td>
<td></td>
<td>culture</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The principal open-ended or “self-diversifying” systems.

By combining finite sets of particulate elements music and language create composite patterns without limit. The role of phonemes, and their higher order combinations into morphemes, in the “duality of patterning” of human language for this purpose is well known (Hockett 1960). Music derives its particulate elements from a radical reduction in the degrees of freedom available to vocal or instrumental sound production by discretizing two continua, those of pitch and duration, to yield musical notes with determinate pitch and — in all rhythmic music — discrete durations (based upon isochronous discretization of time: Merker 2002, 2014a; Merker et al. 2009). The result is the variety of musical notes that serve as grist for the combinatorial mill of musical patterns of boundless variety.

It is in their capacity of Humbold systems that music and language are most fruitfully compared and contrasted. It may even be that there is a parallelism between the two “natural” and the two “cultural” Humbold systems with respect to their genesis. Chemistry obviously is a necessary though not sufficient prerequisite and antecedent for the genetics of life forms. Similarly music in the form of human song may be a necessary but not sufficient prerequisite and antecedent for human language (see further Theses 2, 3, and 4). We might, in other words, have been singing apes before we became talking humans, a possibility that will receive further notice in what follows.
Thesis 2: Genesis

Natural selection is but one — and possibly the least important one — of three agencies likely to have driven the genesis of music and language. The second of these is sexual selection, in accordance with Zahavi’s “handicap principle”, particularly as elaborated in the “developmental stress hypothesis” by Hasselquist. The third is the “learner bottleneck” in cultural transmission across generations, whose powers suffice to transform random strings into a shared grammar without either Darwinian selection or reinforcement of outcomes (Kirby 1998; Kirby et al. 2008 and references therein).

Annotation

The powers of human language to communicate information are so formidable that we all too readily assume that this utility itself drove the evolution of language by natural selection in our line (for which see Pinker & Bloom 1995). If so, one would expect that some of the other large-brained animals on earth would have evolved some version of it as well. Yet among the 15000 species of birds and mammals (Clements 2007; Jones & Safi 2011) — the two taxonomic groups that contain all the world’s large-brained species — there is not a single species with language except our own (Maynard Smith and Szathmáry 1995). This indicates that language is something that simply does not tend to evolve in the animal kingdom.

Any theory of language origin must account for this reluctance of biological evolution to produce language, despite its utility once in hand. There are few ways to escape this quandary except to assume that to arrive at language a species or lineage must come into possession of a number of independent traits or adaptations each of which is rare considered by itself, and all of which together are needed as prerequisites for or stepping stones to language. In such a situation, multiplicative probabilities determine outcomes, and since each trait is rare, their small probabilities multiplied might shrink the product to the infinitesimal level needed to ensure the empirical end result: a single species with language among the millions of animal species on earth. A gapless and biologically plausible path to human language traversing such rapidly dwindling probabilities has been charted in two previous publications of mine (Merker and Okanoya 2007; Merker 2012).

According to the scenario presented there, the expressive powers of our early Homo ancestors did not evolve to serve the communication of information, but as means to impress potential mates and rivals by elaborately structured but unsemanticized vocal displays based on vocal production learning. They evolved, in other words, through sexual selection, in keeping with the handicap principle and the “developmental stress hypothesis” for the evolution of complex learned song in many songbirds and some mammals (Hasselquist et al. 1996; Nowicki et al. 2002; Zahavi and Zahavi 1997). Species thus equipped with the specialized telencephalic mechanism of vocal production learning (see Thesis 4) become carriers of cultural traditions of complex learned song.
Once a species maintains a complex, learned song tradition, the process of intergenerational transmission itself begins to shape and transform the patterns of the transmitted vocal “lore”. When unconstrained by a species-specific song template, as in vocal mimics, this process is even capable of transforming an initial state in which songstrings of arbitrary patterning are randomly distributed across individuals into a state in which they all share a formal grammar. Yet contrary to what one might expect, this process does not require any form of Darwinian selection or reinforcement of outcomes, as shown in large-scale computer simulations of populations of learning agents subject to such intergenerational transmission (Kirby 1998, 2000, 2001, 2002).

The key to this apparent magic is the obligatory dependence of the process of intergenerational transmission on the sparse sampling of the vocal lore available to the individual learner in a population of culture-carriers. The process through which utterances are filtered through this “learner bottleneck” creates competition among vocal patterns for access to subsequent generations. In this competition simple statistics of intergenerational transmission lead more general forms to outlast less general ones in the population over time, eventually assuring convergence on a stable and efficient grammar (see Kirby 2000 for details).

Thus, an aspect of the so called “poverty of the stimulus”, which Chomsky employed in arguments for the necessity of innate syntactic structure (Chomsky 1980, see further Thesis 5), turns out to be a key factor in the emergence of syntax from syntax-free songstrings by learning alone (Smith 2003; Zuidema 2003). It amounts, in other words, to a process of data compression, as efficient as it is slow, requiring hundreds and thousands of generations to do its transformative work. That transformative work takes place unbeknownst to the learners, without active agency on their part, as a glacially slow transgenerational effect of their limited learning capacity, a process sustained by the display function of song, free of the necessity to communicate meaning. Further aspects of this process are briefly touched upon in the annotations to Theses 3, 4, and 5, and in Merker et al. 2015.

**Thesis 3: Paths and precursors**

There is no natural path from the semantics of animal calls to the syntax of human language, but there is a natural path from the syntax of learned animal song to the semantics of human language (Merker 2012). Thus, our path to language must have traversed a stage of learned, unsemanticized, and complex song.

**Annotation**

Animal call systems tend to be innately based, and to feature tight one-to-one coupling between sound gesture and “meaning” (or “functional reference”, see Marler 2004). They are capable of signaling both the varied emotional states of the
animal calling (typically with graded intensity reflecting emotional intensity) and of carrying “functional reference” to external social or environmental circumstances (as in alarm calls: Seyfarth et al. 1980, or food calls: Marler et al. 1992).

Animal calls, being compact meaning-carriers, would accordingly seem ideally placed to serve as “first words” of an open-ended particulate signalling system that conveys new meanings by combining pre-existing call types by syntactic rules, yielding a proto-language. They are, however, inherently incapable of serving in that role.

The problem is this: in order to function as elements of a combinatorial system for conveying meaning, vocal gestures must possess a modicum of neutrality and independence with respect to the basic emotional-motivational forces animating an animal’s behaviour. Consider an animal whose innate repertoire includes two conspicuous calls, one meaning “food”, normally given when the animal finds food, the other meaning “fear”, an alarm call normally given when the animal is frightened, as by a predator. Why not combine these calls to generate four compound meanings from the two, as follows: the alarm call followed by the food call would signal “fear-food”, i.e. “fear-with-food”, meaning “poisonous food”, while the food call followed by the alarm call would mean “food-fear”, i.e. “food-with-fear”, and hence “prey”?

But in order to do so, the animal would have to use a call charged with positive valence (finding food) to signal something with strong negative valence (poison), and also to use a call inherently linked to its own state of fear (as evoked by a predator) to signal something the animal has no reason to fear (its prey). An inherent feature of calls — their strong unitary referential loading, on a typically innate and emotionally charged basis — prevents the animal from doing so. Thus the very traits which make calls so economical and efficient as components of an animal call system erect a functional barrier to the adoption of call combinations as a means to multiply meanings. Given this intrinsic constraint, animals in the state of nature are unlikely to develop call combinatorics much beyond the level documented for the referential call combinations of Campbell’s monkeys (Ouattara et al. 2009).

By contrast, a path from the syntax of animal song to the semantics of human language is encumbered by no such difficulties. Assume a species that has a large repertoire of learned culturally transmitted songstrings composed of sets of discrete elements, as in some songbirds with vocal learning. Unlike such birds, this hypothetical species produces song year round and in the course of all its various activities (a biologically plausible setting for such rare behaviour — a rarity that is welcome given a need for diminishing probabilities — is found in Merker 2012).

Under such circumstances the “learner bottleneck” will segregate songstring usage by behavioral and situational context over generations, saddling songstrings with differential contextual associations. Some songs are sung during hunting, others in gatherings around the camp fire, and so on, extended across the
full spectrum of the species’ activities. In acquiring these passive contextual associations songstrings have, in other words, undergone unplanned and implicit semanticization.

The fact that under these circumstances songstrings will contain various sub-strings that occur scattered haphazardly across strings associated with different contexts provides a mechanism by which the learner bottleneck will associate each such sub-strings with features shared by those contexts in which it occurs. This would result in a glacial process of differential semantic allocation of sub-strings to more general contextual associations over generations, in accordance with principles first enunciated independently by Alison Wray (Wray 1998) and Simon Kirby (1998).

This process, which requires no other operations than those of generalization and segmentation (Hurford 2000; Okanoya and Merker 2007), would produce ever more abstract contextual associations for ever smaller sub-string units, albeit still not in order to communicate information, but purely for display purposes. How this system of highly differentiated contextual songstring associations, which so far are strictly and entirely tacit or implicit, ends up being deployed as a system for communicating information — a language — is outlined in Merker 2012.

In this process there are no “first words” but only long and complex “first (holistic) sentences” (the semanticized songstrings) carrying specific (and initially rather concrete) situational reference on a holistic, unanalyzed basis. Meaningful words appear only at asymptote of the abstraction process by which sub-strings come to be associated with shared features of disparate contexts. That is, words would emerge only in the final phases of glossogenesis, at which point the logic of grammaticalization processes would also become historically active (Bybee 2006; DeLancey 1993; Lehmann 1995; Merker and Okanoya 2007). Incidentally, this process helps explain the historical pattern according to which languages at earlier historical stages often exhibit a more complex grammar than they do today, a result altogether unexpected if language is assumed to originate by gradual accretion from simple to complex.

A special problem in the conversion from a system of song for display purposes to a system of language for conveying information is the fact that the attempt to use semanticized songstrings in “displacement mode”, i.e. to refer to circumstances other than those present, would undermine the contextual associations by which songstrings acquire their semanticization. That is the one point at which natural selection will enter glossogenesis, according to the scenario presented in Merker 2012.
Thesis 4: Vocal learning

Humans are the only primate with a capacity for vocal production learning (Janik and Slater 1997; Egnor and Hauser 2004). That capacity is an absolute requirement for both human song and spoken language in that it allows us to shape vocal output to match auditory models. How and why humans alone added vocal learning to the primate brain holds the biological key to the riddle of human music as well as language (Merker 2009, 2012, 2014c; Nottebohm 1976).

Annotation

An unabridgeable requirement for the existence of cultural traditions of learned songstrings — a key stepping stone to language on the present account — is the mechanism of vocal production learning. It gives us the capacity to reproduce, through the voice, patterns of sound first received by ear (i.e. not already in the innate repertoire of calls). That capacity is absent from all non-human primates, our closest living relatives among the apes included (Janik and Slater 1997; see also Egnor and Hauser 2004).

It is by means of this “watershed adaptation”, added to the primate brain at or after our ancestral split from our common ancestor with chimpanzees, that we acquire every song we know how to sing and every word and phrase we know how to pronounce. Human song and speech did not evolve, in other words, by way of refinements of the “primate communication system”, as is often assumed (see, for example, Ackermann et al. 2014, and commentary by Merker 2014c). Rather, they are products of a mechanism evolved de novo in our line, and distantly shared with numerous species of songlearning songbirds and a few mammals.

Without primate homologs, convergently evolved vocal learning in birds and mammals is the sole source of comparative insight into the biological significance and origin of this trait of ours. The “developmental stress hypothesis” (a sophisticated descendant of Zahavi’s handicap principle) provides the so far most comprehensive biological framework for understanding the function of elaborate vocal production in birds with vocal learning (Hasselquist et al 1996; Nowicki et al. 2002). It also helps answer outstanding questions regarding our own path to vocal learning (Merker 2012).

Evolved through sexual selection for elaborate vocal display, this capacity for vocal production learning provides the key enabling mechanism that brings the data compression powers of the “learner bottleneck” to bear on culturally transmitted vocal lore, as in human song and language (see annotation to Thesis 2).

In vocal learners vocal output is shaped by learning, using feedback of their own voice to match vocal production to heard models (Konishi 2004). This makes the shape of the vocal tract a negligible factor in vocal production, as demonstrated by the ability of bird mimics to duplicate the human voice with their utterly different vocal anatomy (Nottebohm 1976).
Vocal learning requires a lengthy period of unreinforced practice to achieve duplication of heard patterns. This necessitates a motivational mechanism to sustain practice till a match is achieved. This “conformal motive” (Merker 2005), once in place, provides a pivot for generalization to other imitative abilities, and may account for the unique imitative, expressive, and ritual propensities of humans, including their ready acceptance of arbitrary words as the obligatory names of things. It may account, in other words, for the length to which humans, alone among primates, have gone in developing a capacity for “expressive mimesis” (for which see Donald 1991).

**Thesis 5: Biological endowment**

*In a big-brained primate with a capacity for vocal production learning (see Thesis 4), a biological endowment for human language is unlikely to take the form of the innate, universal grammar postulated by Noam Chomsky (Chomsky 1968, 1975, p. 38). Given that the “learner bottleneck” in transgenerational learning traditions has the power to ensure language learnability (Christiansen and Chater 2008; Kirby et al. 2008 and references therein; Merker 2009; Zuidema 2003), the mechanism of vocal production learning that sustains traditions of learned vocal lore constitutes our biological endowment for human language and song (Merker 2009).*

**Annotation**

Noam Chomsky confounded the question of the nature of a human “endowment for language” with the issue of the learnability of language by means of the formal behaviourist learning theory of his time (Merker 2009). He took its unlearnability by those means to imply that we are possessed of a so called universal grammar on an innate basis (Chomsky 1975, *in* Piatelli-Palmarini 1980, p. 111).

One need not follow Chomsky in this first false step into the biology of language in order to recognize the existence of a human “endowment for language”. That endowment is supplied by our telencephalic mechanism dedicated to vocal production learning discussed in Thesis 4. Once this mechanism was in place for purposes of handicap-driven song displays, the learner bottleneck brought its data compression powers to bear on the pattern content of the intergenerationally transmitted vocal lore, with consequences briefly outlined in Theses 2, 3, and 4.

Supported by a conformal motive (Merker 2005) and *de novo* evolution of a direct projection from primary motor cortex to the respiratory and phonatory motor nuclei of the lower brain stem (Brown *et al.* 2008; Okanoya & Merker 2007), vocal learning turns the cerebral territories centred on Wernicke’s and Broca’s areas from their non-language uses in other primates to the service of human language by recruiting them to the generative production and intergenerational transmission of culturally learned vocal lore (for details, see Merker 2009, 2014c; Okanoya & Merker 2007).
To this dedicated cerebral mechanism for vocal learning we owe not only our developmental trajectory for language learning, infant babbling included (Doupé & Kuhl 1999), but our propensity for imitation and ritual culture more generally (Merker 2005), along with a robust selection pressure for encephalization (Merker & Okanoya 2007; Merker 2012). Through vocal learning, and unconstrained by innate so called universal grammar, the historical filter of cultural transmission — which passes only the possible (Zuidema 2003) — continually adapts the actual forms of languages to multiple interacting constraints such as use, utility, learnability and neural resources (Christiansen & Chater 2008), as well as to cultural norms (Everett 2005).

An endowment for human language in the form of vocal production learning, as opposed to so called universal grammar, explains the ineluctably historical nature of human languages, for which grammaticalization theory supplies some of the general laws (Bybee 2006; Lehmann 1995). The principal consequence of this inherently historical nature of human language is the empirically well attested rampant structural diversity exhibited by the world’s languages (Evans & Levinson 2009).

**Thesis 6: Gesture**

The plausibility of a gestural origin of language (Arbib & Rizzolatti 1996; Corballis 1999) can be estimated by the likelihood that a species already in possession of one functioning form of language (manual-visual) would switch to another (vocal-auditory) at the cost of having to develop the entire machinery of vocal production learning de novo in order to do so, and this for benefits such as “freeing the hands” from their encumbrance by gestural language, an encumbrance no greater than that it did not prevent that species from developing gestural language in the first place.

**Annotation**

An absolute prerequisite for language in the vocal-auditory mode is vocal production learning (see Thesis 4). The absence of that capacity in our common ancestor with chimpanzees might seem to make a gestural origin of language an attractive possibility, since it would draw on the far more widely distributed learning capacity for manual dexterity, which our closest primate relatives do possess. The supplemental role of gesture in speech, and the ease with which humans acquire or develop sign language when the auditory-vocal mode is unavailable, has been taken to support a gestural origin of language in humans (Arbib & Rizzolatti 1996; Corballis 1999).

That, however, does not remove the vocal learning difficulty; it only postpones it, because the end state of language evolution is vocal-auditory speech. How would a species already in possession of gestural language implemented as a manual-visual
system of conventional gestures ever switch to vocal language in the absence of a capacity for vocal learning, which is an absolute requirement for human speech?

Obviously, if chimpanzees were vocal production learners and easily acquired a spoken vocabulary, a gestural origin for language would lose all plausibility. But they are not, so our common ancestor with chimpanzees can be assumed to have lacked vocal learning. That lack, which favours a gestural origin, must conveniently be forgotten in order to explain how the original gestural mode of language (manual-visual) was replaced by a radically different mode (vocal-auditory) at reasonable cost.

Assuming an absence of the capacity for vocal learning at that point forces us to assume that the advantages of a switch sufficed to pay the evolutionary cost of developing the capacity for vocal learning de novo. The encumbrance of the hands by gestural language is no greater than that it allowed the evolution of gestural language in the first place, on the gestural theory. Other advantages of speech, such as communicating propositionally when the audience is out of sight, are also marginal, and come with disadvantages as well: vocal language makes both predators and prey privy to the communicative act, even when out of sight.

How encumbered are sign language users, really, compared to speakers, except by the circumstance that the population at large speaks rather than signs? Is whatever they are lacking by having their hands “encumbered” by language of sufficient magnitude to serve as a selection pressure for the evolution of a sophisticated cerebral machinery for vocal production learning, a machinery which up to that point has not made its appearance in any other primate?

The obvious alternative is to assume that human language originated in the auditory-vocal mode, because we had already evolved vocal learning for non-speech purposes before employing it for language. In us, as in most species that possess it, vocal learning is likely to have evolved for learned song, as outlined in Theses 2, 3, 4, and 5. Under such circumstances, a gestural origin of language loses its appeal. Given that the end-state is vocal-auditory, a species already in possession of vocal production learning has no reason to make a detour into gestural language along its path to speech.

This is all the more so since manual gestures do in fact have a natural, if subsidiary, role to play in the vocal learning path to speech. They are needed as auxiliaries in the transition to the use of semanticized songstrings in “displacement mode” (Hockett 1960) on the vocal path to human language, as markers for communicative intent in that mode, as detailed in a previous publication of mine (Merker 2012, p. 244). That subsidiary role is eminently compatible with the way present day speakers supplement spoken language with manual gestures.

Finally, and parenthetically, the acquisition of conventional gestures is inherently imitative. Our unusual imitative capacities compared to our primate relatives likely arose by generalization from our capacity for vocal production learning, as outlined in the annotation to Thesis 4, following Merker (2005). No
other species has, after all, developed a system of conventional manual gestures for language-like communication, even though the manual learning capacity required for it is available, as shown by apes taught sign language in captivity (e.g. Miles 1990). The lack of vocal production learning from which to generalize high-fidelity duplicative behaviour may explain this anomaly, and reinforces the plausibility of a vocal rather than gestural origin of human language.

Thesis 7: Meaning

Language alone “means”, which is to say that it conveys articulated semantic content through a system of multi-level conventional coding rather than by having its patterns in any way resemble their referents (Staal 1989). There is no corresponding device in music, whose patterns do not serve as structural vehicles for conveying another domain of content than themselves, except when they are made to do so (in programmatic music) by means of mimicry or resemblance (Hanslick 1854).

Annotation

Language in the form of speech embodies a bona fide code for information in that it avails itself of a multilevel combinatorics of phonological and lexical elements to perform arbitrary (in the sense of conventional) mappings between the form of utterances and their meaning. An elementary example is the combinatorics of phoneme sequence that assigns distinctive meanings to the tens of thousands of words that make up our lexicon. Thus bord, Tisch, and table are such arbitrary phonological combinations for the self-same type of physical object, in Swedish, German and English, respectively. A higher level example is the grammatical device of sequential ordering of major sentence constituents such as subject, verb and object (for which essentially every possible variant ordering is attested across the world’s languages [Comrie 1983]).

This language code of meaning is so detailed and comprehensive that virtually every difference between strings of phonemes makes a difference in the information conveyed by those strings. By means of this elaborate, multilevel conventional coding arrangement of phonemes and morphemes, sequential patterns of vocally produced sounds become a vehicle for making statements about things that bear not the slightest resemblance to those sound sequences themselves (see the table example above). Thus language allows us to communicate about objects, events, matters of fact, states of the world, ideas, intentions, beliefs and desires, without limit.

Nothing of the kind takes place in human song and music (unless they feature lyrics), however structurally complex they may be. In song (and music) it is the vocal (instrumental) patterns themselves that are the information conveyed, not another domain of content which rides upon those patterns by means of a syntactic
system of conventional coding. Attempts have been made to assimilate music to language by treating music as a “language of emotions” (Spencer 1911). However, the weak sense in which music is capable of evoking or portraying emotions (Koneční 2008, 2015; Koneční et al. 2008; Scherer 2003) or circumstances in the world (storms, battles: Hanslick 1854) is achieved by using the pattern richness of music to dynamically mimic, resemble, or caricature the things to be evoked (say, the mimicking of emotional prosody as a means to portray emotion in Western music [Juslin 2001]).

So even when music attempts to portray — which is far from always the case — it does not mean, it mimics, a distinction easily accommodated within the conceptual apparatus of semiotics. Such mimicking is exactly what language does not do, except in the trivial case of onomatopoeia. There is, however, an aspect of speech that does so. We have a rich repertoire of non-verbal vocal expressiveness, ultimately based upon our innate calls (see Thesis 3). It coexists with the meaning-code of language in speech. That non-verbal expressiveness enters speech by means of its prosody, and is used among other things to give speech its emotional colouring. To confuse that aspect of language with its meaning-carrying coding arrangements would of course be fatal. The two dwell in different qualitative and formal universes and must not be confused.

Thus, the fact that music can be made to suggest or mimic emotion tells us as little about the nature of its generative pattern richness as the facts of speech prosody tell us about the meaning-carrying generative pattern richness of language. The secret of musical pattern richness (see Thesis 1) is best approached by abandoning the notion that it contains a code for a domain of information other than those patterns themselves.

Those patterns expose our perception and imagination to their structural pattern content: their mix of variation and repetition, their symmetries and asymmetries, the evolution of their temporal trajectories with their tensions and resolutions, their intricacy, complexity, or simplicity, the balance of such factors among themselves as the music unfolds in time, and more. It is not to our emotions that this content is addressed in the first place, but to our imagination, as Arthur Schopenhauer, followed by Eduard Hanslick, insisted (Hanslick 1854; Merker 2007; Shopenhauer 1844, vol. 2, p. 447ff).

In so doing music will exhibit greater or lesser elegance, will achieve greater or lesser degrees of well-formedness (see, e.g., Lerdahl & Jackendoff 1983) and — depending on our background of acquired familiarity forged by our listening history — may compel our attention more or less fully, or please us more or less well. In all of these respects the patterns of music define themselves as esthetic objects that address the informational dimension of our cognitive capacities directly, in the manner of auditory analogues of visually presented arabesques (Hanslick 1854).
This by no means relegates music to an abstract domain of passively contemplative connoisseurship. Some of its patterns — say of rhythmic music meant to accompany dancing — access a presumably species-specific predisposition for bodily entrainment to isochronous auditory patterns, and help optimize such entrainment (Merker et al. 2009; Merker 2014a). Its central role in youth and popular culture suffices to dispel any overly formalist notion of its nature. It is also at the farther reaches of this “informational dimension of our cognitive capacities” that music exerts its own genuinely emotional effects.

Just as an elegantly crafted arabesque differs in its impact on our imagination and aesthetic sensibilities from that of a crude and clumsy doodle, so musical patterns have analogous differential impact on our sensibilities by their pattern characteristics alone. Where an arabesque must “look good” to capture our imagination, musical patterns must “sound good”, the more so the greater their impact. So much so that those that sound exceptionally good tend to impress, even to the point of inspiring awe and a sense of the sublime, emotions so powerful as to drive tears to our eyes and chills up our spines (Gabrielsson 2011; Konečni 2005, 2011; Merker 2014b).

That is where the notion of an emotional impact of music has its true point of support: not in any analogy to the meaning encoded in language, nor in any kinship with the biology of basic emotions, but in the basic biology of the Zahavian handicap principle (Zahavi 1975). Where handicaps take the form of esthetic displays — from the peacock’s tail to the vocal artistry of pied butcherbirds — appraisal mechanisms for judging their quality must be in place, typically in the medium of a cognitive dimension spanning from boredom, via interest/curiosity, to being impressed, with awe and a sense of sublimity at its high end (Konečni 2005, 2011; Merker 2012, 2014b).

The elaboration of Zahavi’s handicap principle in the developmental stress hypothesis to account for complexity and size of birdsong repertoires by Hasselquist and colleagues (Hasselquist et al. 1996; Nowicki et al. 2002; see Theses 2 and 4) provides an eminently plausible interpretive framework for the nature and function of human song and music as well. It dispels the appearance of frivolity encumbering our expenditure of effort and resources on acquiring and producing the pattern richness of human song and music. Our musical developmental trajectory does not equip us with a language-like device for conveying referential meaning (emotional or otherwise). Rather, by exact analogy to the case of learned birdsong, it gives us a means to display command and mastery of a trove of culturally patterned and transmitted lore. Such command and mastery serves not only as a badge of competence in the culture, but as a certificate of the phenotypic traits needed to achieve that competence (see further Merker 2012, 2014b).
Bibliographical References


