



# University of HUDDERSFIELD

## University of Huddersfield Repository

Jan, Steven

From holism to compositionality: memes and the evolution of segmentation, syntax, and signification in music and language

### Original Citation

Jan, Steven (2015) From holism to compositionality: memes and the evolution of segmentation, syntax, and signification in music and language. *Language and Cognition*. ISSN 1866-9808

This version is available at <http://eprints.hud.ac.uk/id/eprint/23707/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: [E.mailbox@hud.ac.uk](mailto:E.mailbox@hud.ac.uk).

<http://eprints.hud.ac.uk/>

**FROM HOLISM TO COMPOSITIONALITY: MEMES AND THE EVOLUTION OF  
SEGMENTATION, SYNTAX AND SIGNIFICATION IN MUSIC AND LANGUAGE**

---

**Author Details**

Dr Steven Jan

Department of Music and Drama

Creative Arts Building

University of Huddersfield

Queensgate

Huddersfield HD1 3DH

United Kingdom

E: [s.b.jan@hud.ac.uk](mailto:s.b.jan@hud.ac.uk)

T: +44 (0) 1484 472 143

W: <http://www.hud.ac.uk/ourstaff/profile/index.php?staffuid=smussbj>

## **Abstract**

Steven Mithen argues that language evolved from an antecedent he terms “Hmmmmm, [meaning it was] Holistic, manipulative, multi-modal, musical and mimetic”. Owing to certain innate and learned factors, a capacity for segmentation and cross-stream mapping in early *Homo sapiens* broke the continuous line of Hmmmmm, creating discrete replicated units which, with the initial support of Hmmmmm, eventually became the semantically freighted words of modern language. That which remained after what was a bifurcation of Hmmmmm arguably survived as music, existing as a sound stream segmented into discrete units, although one without the explicit and relatively fixed semantic content of language. All three types of utterance – the parent Hmmmmm, language and music – are amenable to a memetic interpretation which applies Universal Darwinism to what are understood as language and musical memes. On the basis of Peter Carruthers’ distinction between “cognitivism” and “communicativism” in language, and William Calvin’s theories of cortical information encoding, a framework is hypothesized for the semantic and syntactic associations between, on the one hand, the sonic patterns of language memes (“lexemes”) and of musical memes (“musemes”) and, on the other hand, “mentalese” conceptual structures, in Chomsky’s “Logical Form” (LF).

## **Keywords**

Hmmmmm, music, language, meme, proteme, museme, lexeme, compositional, cognitivism, communicativism.

---

## **1. Introduction: The Evolution of Music and Language from “HmMMMM”**

Theorization on the origin of language has been enriched in the last decade or so by considering the issue in conjunction with discussion of the origin of music (Patel, 2008, Chapter 7).<sup>1</sup> This follows decades of separating their treatment,<sup>2</sup> a strategy which often goes hand in hand with seeing language as prior to music and which, at its most extreme, leads to the view that music is a mere by-product, frivolous and indulgent, of linguistic ability (Pinker, 1997, p. 528). Recent research has considered more systematically views first expressed in the eighteenth and nineteenth centuries which saw the two domains as intimately connected. Such integrated conceptions tend to view music not as a successor to but as a precursor of language. While Darwin’s statement that “the progenitors of man probably uttered musical tones before they had acquired the power of articulate speech; and that consequently, when the voice is used under any strong emotion, it tends to assume, through the principle of association, a musical character” (in Gamble, 2012, p. 83) is perhaps the most well known of these, the dependence of language upon music was recognized not only by Otto Jespersen after Darwin’s time but, in the eighteenth century, by Jean-Jacques Rousseau (Mithen, 2006, p. 2) and Johann Gottfried Herder (Bohlman, 2002, p. 39).

Of contemporary theorists in this tradition, Steven Mithen’s thesis is perhaps the most convincingly argued. He holds that music and language shared a common ancestor, a form of primal song, which gradually bifurcated into the two modern forms, with music retaining the melodiousness of the original protolanguage while losing its (limited) referential capacity; and language acquiring stable semantic content while losing many of the more overtly musical inflexions of its parent (Mithen, 2006). In this way, the “singing Neanderthal” gave way to the speaking human; indeed, the evolutionary utility of developed language in *Homo sapiens* may explain in part the extinction of the Neanderthals and our survival.

Mithen sees hominin protolanguage in a different sense to Derek Bickerton, the latter believing it was made up of “words, with limited, if any, grammar” (Mithen, 2006, p. 3; Bickerton, 2003). Mithen terms his primal song “HmMMMM, [meaning it was] Holistic, manipulative, multi-modal, musical and mimetic” (Mithen, 2006, pp. 138, 172). That is (according to Mithen’s speculations), its component gestures could not, *contra* Bickerton, be decomposed into individual meaning-units (protowords), but were to be understood as constituting a single unified message; it was designed to affect and mediate the thoughts and behaviour of others, often to the advantage of the utterer; it drew not only upon sonic elements, but also upon physical gestures and movements, actions and facial expressions; it was what we today might easily regard as a form of music, in that it consisted of interconnected melodic phrases which combined pitch, rhythm and, presumably, dynamics and timbre; and it was often imitative of the sounds of the world of the utterer – those of the birds, animals and other natural phenomena which constituted the environment of the hominin species which utilized it. Mithen argues that HmMMMM was employed (to list the hominin line in hypothesized order of appearance) by *Homo ergaster*, *Homo erectus*, *Homo heidelbergensis*, *Homo neanderthalensis* and early *Homo sapiens* (Mithen, 2006, p. 7, Figure 1; see also Foley, 2012).

If HmMMMM constituted a form of holistic communication, then “modern” language – that is, the broadly word-based, syntax-governed form of communication which began to evolve in *Homo sapiens* after 200,000 years ago (Mithen, 2006, p. 257) – is by contrast “compositional”. That is, it is made up of relatively discrete sonic units which may be recombined (often recursively/hierarchically) according to the principles of some grammatical system, in order to assemble a near-infinity of potential utterances, thereby vastly exceeding the flexibility and communicative power of holistic forms of communication. While each sonic unit in compositional language has a fairly stable semantic content, this may change according to the grammatical function of the unit within the utterance, as exemplified by Truss’s celebrated amphibology “eats[,] shoots and leaves” (Truss, 2003). That the Neanderthals never learned to shoot, despite eating shoots and leaves, might be a consequence of a lack of the expansion in thought and invention facilitated in *Homo sapiens* by compositional language.

While Mithen’s account is painstakingly outlined and convincingly supported, it can be argued that it lacks a consistently Darwinian focus. That is, while it incorporates Darwinism in its account of

the genetic basis of language – in its consideration of such interconnected aspects as bipedalism, the evolution of the vocal tract, and sexual selection (Mithen, 2006, pp. 139ff., 146, 176ff.) (for more, see Section 2) – it does not complement this by a consideration of Darwinism’s operation in the cultural dimension. It does not, in other words, embrace what Richard Dawkins terms “Universal Darwinism” (Dawkins, 1983) in accepting the influence of the evolutionary algorithm (Dennett, 1995, p. 343) on the extra-genetic (cultural) dimension. In this sense Mithen does not offer a fully *co-evolutionary* account of language (Durham, 1991) which recognizes the ways in which the evolutionary advantages of each replicator, natural (gene) and cultural (meme), interact.

This article attempts to redress this imbalance by reconceiving the process of language evolution Mithen outlines in Universal-Darwinian terms, arguing that Mithen’s “mimetic” can be replaced by Dawkins’ “memetic” (Dawkins, 1989). That is, the self-interested replicated particle offers a means of arriving at a unified cultural-Darwinian conception of language and music which understands their similarities and differences as a consequence of their aetiology. Moreover, it allows a mediation between their phonetic, syntactic and semantic dimensions, and their neurological and psychological bases.

Taking this position, and by way of a conceptual overview, I attempt to address the following issues:

1. The hypothesized co-evolutionary origins of music and language, whereby Hmmmmm bifurcated into language and music.
2. The memetic similarities between language and music, understood in terms of segmentation and replication.
3. The relationship between language, thought and consciousness, and to what extent theorization on this issue might be extrapolated to music.
4. The relationships between neurological mechanisms of language and music encoding, including hemispheric localization.
5. The possibility, on the basis of their hypothesized common evolutionary origin, of inferring referential content in music by extrapolation from the referential-semiotic mechanisms underpinning language.
6. The analogous possibility of inferring grammatical content in music by extrapolation from the grammatical mechanisms underpinning language.

Section 2 explores how Hmmmmm may have become segmented into discrete units capable of being subject to the Darwinian algorithm, and review how this has been simulated computationally. Section 3 considers some of Peter Carruthers’ ideas on the relationship between language and thought and asks to what extent the linguistic descendant of Hmmmmm was a means for reflecting thought and to what extent it actually constituted thought. Section 4 explores William Calvin’s hypothesis for the mechanisms underpinning information encoding in the brain, arguing that the component units of music and language are implemented in cortex in configurationally and functionally similar ways, in accordance with their hypothesized common evolutionary origins. Section 5 attempts a memetic synthesis of Mithen’s, Carruthers’ and Calvin’s ideas, arguing how syntax and semantics in music and language might follow broadly similar mechanisms. Specifically, it outlines a potential homology between “mentalese” (brain-language, inaccessible directly to consciousness) and language-sounds that might be mirrored in a parallel alignment between mentalese and music-sounds. Section 6 reviews the arguments advanced and suggests some ways in which they might be extended.

## **2. A Memetic View of the Evolution of Hmmmmm**

I consider here how Mithen’s hypothesized Hmmmmm evolved into compositional language by means of various gestalt-psychological pressures and the Universal-Darwinian/memetic processes

these and other factors engendered. Before beginning this, it is worth briefly summarizing the associated biological-evolutionary changes which formed the context, and to some extent the motivation, for music-linguistic evolution. In broadly sequential order, the salient events linking the last common ancestor of *Homo sapiens* and *Pan troglodytes* (chimpanzee) to *Homo ergaster* appear to have been as follows:

1. Movement from a predominantly arboreal lifestyle to one of savannah dwelling (starting perhaps with *Australopithecus afarensis* c. 2 million years ago (MYA)) led to the evolution of bipedalism (Mithen, 2006, pp. 144–145). One consequence of the various anatomical and physiological changes impelled by bipedalism was the lowering of the larynx, the consequent increase in pharyngeal space, and a resultant augmentation of vocal range and control (Clegg, 2012).
2. Savannah dwelling implies a greater tendency to communal living, for mutual protection and maximization of resources. For males, this necessitated increased cooperation in hunting (given the increasingly carnivorous diet implied by this lifestyle); for females, it implied increased cooperation in foraging and infant-rearing, the latter including “grandmothering” – the co-opting of post-menopausal females in support of food-gathering mothers. The latter developments may partly have been a consequence of the constraints of birth-size imposed by bipedalism, hominin infants requiring several years of nurture before they become independent of their parents (Mithen, 2006, pp. 185–186).
3. Starting with *Homo ergaster* at c.1.8 MYA, the modern human male : female size ratio of c. 1.2 : 2 became established (perhaps because males had reached the limits of their size and females had nearly caught up with them). High levels of sexual dimorphism correlate with polygyny (the sexually selected outcome of males competing violently with each other for female attention and/or forcing themselves upon females); and low levels of sexual dimorphism (in *Homo ergaster* and subsequent species) correlate with (sometimes serial) monogamy, with males providing food and care for their mates and their dependent children (Mithen, 2006, pp. 182–187).
4. Male-to-male competition continued after the levelling off of sexual dimorphism, but was less focused upon male force and more upon female choice, given females’ broadly equivalent size to males and their ability to draw upon defensive alliances with other females. Such male-to-male competition involved convincing females by means other than physical force that they possessed good genes, and singing and dancing appear prime candidates for this (Mithen, 2006, p. 187). In this sense, singing and dancing become currencies implicated in sexual selection.
5. The mutual grooming common in primates intended to foster networks of social relationships became more difficult as social groups increased in size and was, according to Aiello and Dunbar (1993), supplanted by “vocal grooming”. Here, Hmmm allowed one individual efficiently to interact with multiple recipients, maximizing that individual’s pay-off in the form of reciprocal attention.

## **2.1. Segmentation and Cross-Stream Mapping**

Mithen argues that the factor which drove the evolution from Hmmm to compositional language was segmentation – “the process whereby humans began to break up holistic phrases into separate units, each of which had its own referential meaning and [which] could then be recombined with units from other utterances to create an infinite array of new utterances” (Mithen, 2006, p. 253). It is worth first noting that while a word might appear discrete and self-contained on the printed page – the surrounding characters’ worth of whitespace affording the necessary gestalt grouping clue to demarcate its group of letters from other groups – in spoken language a word is usually part of a continual, unbroken sound-stream, and its isolation relies upon a number of segmentational factors.

One might call these separate units “protemes”, in order to signify that they were the evolutionary precursors to both “musemes” and “lexemes”. I use the latter two terms here in a slightly different sense to their normal usage, employing the suffix “-eme” to denote derivation from “meme”. Thus,

museme is a contraction of “musical meme” (and not, in Tagg’s sense, and after Seeger, “a complete, independent unit of music-logical form or mood” (in Tagg, 1999, p. 32)); and a lexeme is a replicated verbal unit (and not (just) “a unit of lexical meaning, which exists regardless of any inflectional endings it may have or the number of words it may contain” (Crystal, 2003, p. 118)). In all three types (in my specific senses of the second and third types), memetic replication is key to their ontology.<sup>3</sup>

There are a number of interrelated processes by means of which segmentation of HmMMMM into protemes could have occurred. The first, according to Alison Wray (Wray, 1998), is the result of “the recognition of chance associations between the phonetic segments of the holistic utterance and the objects or events to which they related. Once recognized, these associations might then have been used in a referential fashion to create new, compositional phrases” (Mithen, 2006, p. 253). While certainly a credible hypothesis, it appears to be predicated upon the existence of another, arguably prior, process to make it function: the presence of some innate psychological tendency which perceives (and imposes) segmentation boundaries at certain points of an ostensibly holistic sound stream, in order to create the “phonetic segments” of which Wray speaks.

Generally considered under the rubric of gestalt psychology, it is well understood that certain factors in a sound stream tend to impose a segmentation boundary (Deutsch, 1999), breaking it up into discrete units. As Eugene Narmour argues, “unlike the notoriously interpretive, holistically super-summative, top-down Gestalt laws of ‘good’ continuation, ‘good’ figure, and ‘best’ organization ... the [bottom-up] Gestalt laws of similarity, proximity, and common direction are measurable, formalizable, and thus open to empirical testing” (Narmour, 1989, p. 47). Thus, where similarity becomes difference, where proximity becomes distance, and where common direction becomes a change in (pitch) direction, a segmentation boundary is likely to be perceived. Moreover, this factor combines with the limits of short-term memory (STM) to impose a limit on the size of the “chunks” which lie in between segmentation boundaries (Snyder, 2009, p. 108). In Miller’s well known formulation, it is “seven, plus or minus two” units (G. A. Miller, 1956); for Temperley, it is “roughly 8 notes” (Temperley, 2001, p. 69).

An additional, supporting process at play in segmentation and meaning-assignment is what I term “coindexation-determined segmentation” (Jan, 2011, sec. 4.1.2), a form of cross-stream mapping. Here, an overlap between two sound streams (one stored in memory, the other heard in real time) imposes a segmentation boundary, provided it is not strongly contradicted by gestalt forces, at the start and end of the “overlapping” segment and affords the common pitches greater perceptual-cognitive salience than they would otherwise have possessed. In other words, as Calvin argues, “that which is copied may serve to define the pattern” (Calvin, 1996, p. 21). Coindexation-determined segmentation might be regarded as culturally (as opposed to genetically) mediated, and an example of the operation of what Narmour terms “extraopus style” (Narmour, 1990, pp. 35–38). As such, it is likely to be more malleable – and therefore more evolutionarily variable (even dialect-mediating (Meyer, 1996, p. 23)) – than genetically mediated segmentation.

Given the presence of gestalt grouping, STM chunking and coindexation-determined segmentation, Wray’s “recognition of chance associations between the phonetic segments of the holistic utterance and the objects or events to which they related” is eminently feasible. Assuming the alignment of these various processes, overlapping, gestalt-demarcated segments would have acquired a distinct identity, and the association with specific “objects or events” would have become ever more firmly established.

Such associations may initially have been “iconic” (segmented verbal chunks acting mimetically as “signs that are motivated by similarity” to that with which they come to be associated) (and so not strictly “chance associations”); but later they may have become “indexical” (chunks “motivated by contiguity or co-occurrence” with that with which they come to be associated) (thus more properly “chance associations”) (Tolbert, 2001, p. 88; see also Cross & Tolbert, 2009, p. 25). On the grounds that Deacon argues that “the criterial attribute of human symbolic thought is arbitrary reference displaced from its immediate context, and that displacement is a function of the hierarchical structure of symbolic thought” (in Tolbert, 2001, p. 88), one might assume the chronological priority of the iconic over the indexical.

Clearly any innately driven segmentation relies upon genetic factors for its implementation, and it is necessary to appreciate that those factors which are present in modern *Homo sapiens* may not necessarily have been in place in earlier hominin species. In this case, earlier hominins would not have possessed the capacity to hear a holistic utterance as anything other than an undifferentiated sonic continuity. Mithen considers the role the FOXP2 gene (Carroll, 2003) may play in language. He cautions that “FOXP2 is not *the* gene for grammar, let alone for language. There must be a great many genes involved in providing the capacity for language, many of which are likely to play multiple roles in the development of an individual” (Mithen, 2006, p. 250; his emphasis). Nevertheless, he hypothesizes that “[p]erhaps the process of segmentation was dependent upon this gene in some manner that has yet to be discovered” (Mithen, 2006, p. 258); and he notes that studies suggest that those with a faulty version of this gene (such as the “KE” family, to whom he refers) encounter “difficulties ... with the segmentation of what sound to them like holistic utterances” (Mithen, 2006, p. 258). The FOXP2 gene could therefore be hypothesized indirectly to underpin segmentation, in that it might subserve certain gestalt grouping principles in perception, it might mediate the length-constraints of STM, and it might support the recognition of similarity in cross-stream mapping.

## 2.2. Computer Simulation of Linguistic Evolution

Discussing Wray’s “associations between the phonetic segments ... and the objects or events”, Mithen argues that a listener “infers some form of non-random behaviour in a [speaker] indicating a recurrent association between a symbol string [protome] and a meaning, and then uses this association to produce its own utterances, which are now genuinely non-random” (Mithen, 2006, p. 256). Mithen’s remarks in fact apply specifically to computer simulations of this hypothesized process by Simon Kirby and his colleagues (Kirby, 2001, 2007, 2013; Scott-Phillips & Kirby, 2010). These are motivated by the desire to understand how compositionality evolved in language, using the computer to replicate in minutes processes which occurred over many thousands of years and which are, of course, not directly accessible to us. Such simulations suggest that Mithen’s hypothesis, after Wray, for the evolution of language from Hmmmm is feasible in principle.

In one “iterated learning model” (ILM) study (Kirby, 2001), Kirby used agent-based simulation to model the transmission of language between an adult agent and a learner agent. He made a distinction between meaning (expressed here simply as a two-component pattern  $a, b$ , each component of which had a value between 0 and 5 (e.g.,  $a_0, b_3$ )) and signal (here a character string drawn from the letters a–z) (Kirby, 2001, p. 103). After the first 50 utterances by the adult, it became evident that a form of protolanguage had evolved (Kirby, 2001, p. 105). By a later stage of the simulation the system had converged on a fully compositional language (Kirby, 2001, p. 106) in which meaning and signal had aligned closely under the aegis of a controlling grammar. Further refinement of the system allowed it to generate “stable irregularity”, of the type common in natural languages where, for example, some of the most common verbs are highly but stably irregular (Kirby, 2001, p. 107).

Kirby sees this as a vindication of Wray’s “associations...” hypothesis, arguing (apropos a later simulation) that

similarities between strings that by chance correspond to similarities between their associated meanings are being picked up by the learning algorithms that are sensitive to such substructure. Even if the occurrences of such correspondences are rare, they are amplified by the iterated learning process. A holistic mapping between a single meaning and a single string will only be transmitted if that particular meaning is observed by a learner. A mapping between a sub-part of a meaning and a [segmented, protemic] sub-string on the other hand will be provided with an opportunity for transmission every time *any meaning is observed that shares that sub-part*. Because of this differential in the chance of successful transmission, these compositional correspondences tend to snowball until the entire language consists of an interlocking system of [meaning-protome] regularities (Kirby, 2013, pp. 129–130; his emphasis).

Kirby is at pains to stress that his (2001) system is focused “less on the way in which we as a species have adapted to the task of using language and more on the ways in which languages adapt to being better passed on by us” (Kirby, 2001, p. 110). Languages have to adapt (towards greater

compositionality) because “[h]olistic languages *cannot* be reliably transmitted in the presence of a [learner] bottleneck . . . , since generalisation to unseen examples cannot be reliable” (Kirby, 2013, p. 129; his emphasis). Thus, in his model “there is no natural selection; agents do not adapt, but rather we can see the process of transmission in the ILM as imposing a cultural linguistic [i.e., memetic] selection on features of the language that the agents use” (Kirby, 2001, p. 108).<sup>4</sup> As Merker elegantly summarizes the process, as it occurs both naturally and electronically,

[t]he [song] repertoire . . . is launched on a process of progressive string-context assortative and hierarchical decomposition from holistic strings *downwards*. Taking place as an unintended side effect of intergenerational transmission through the learner bottleneck, the process is entirely passive and automatic, and takes place [initially] for no reason of instrumental utility whatsoever (Merker, 2012, pp. 241–242; his emphasis).

### **2.3. Echoes of HmMMM in the Modern World**

Might we be able to reconstruct HmMMM, hearing again the sounds which daily echoed around the locations of hominin communities? At first thought, this might seem impossible, because the hypothesized bifurcation between music and language occurred, as noted, some 200,000 years ago and the essence of the parent HmMMM (while surviving vestigially in music and language) might be assumed to have been lost through this division. But it might be possible to find a “living fossil” of HmMMM, analogous to the Coelacanth once long to have become extinct in the Late Cretaceous period but discovered alive in 1938. Three candidates for persistent HmMMM appear to exist, which I address in order of their increasing similarity to music.

The first is the phenomenon of tone languages, wherein meaning is communicated in part by the production of words at specific pitches, either fixed (‘level tones’) or mobile (‘contour tones’) (Patel, 2008, p. 39). While over half the world’s languages are tonal (including most African and south east Asian languages), only a very small minority use the apparent maximum of five level tones (Patel, 2008, pp. 40, 41). The Amazonian Ticuna language appears a strong candidate for the one most proximate to HmMMM, in having five level tones and seven ‘glides’ from one pitch to another (Patel, 2008, p. 42, Figure 2.12).

The second candidate seems clear: IDS (infant-directed speech), or “motherese”, a form of speech-song-gesture communication ubiquitous in human cultures (Morley, 2012, p. 126). Mithen argues that “when we hear mothers, fathers, siblings and others ‘talking’ to babies, are we perhaps hearing the closest thing to ‘HmMMM’ that we can find in the world today?” (Mithen, 2006, p. 275).

As a third candidate, and despite his assertion that IDS is the most likely contender for the persistence of HmMMM, Mithen later goes on to offer an alternative, in his view stronger, candidate, in the form of the mantras of eastern religion. Mithen suggests that, “[a]s relatively fixed expressions passed from generation to generation, [mantras] are, perhaps, even closer than IDS to the type of ‘HmMMM’ utterances of our human ancestors” (Mithen, 2006, p. 277). Mantras exist in many different forms according to the specific religious tradition from which they spring – Hinduism, Sikhism, Buddhism, or Jainism. But many align closely with the hypothesized attributes of HmMMM, in that they exist as melodic-melismatic elaborations of one or more syllables. According to Mithen, “[t]he philosopher Franz [*sic*] Staal . . . concluded that these lengthy speech acts lack any meaning or grammatical structure, and are further distinguished from language by their musical nature” (Mithen, 2006, p. 277).<sup>5</sup>

Despite Mithen’s regarding mantras as the strongest candidate for residual HmMMM, IDS is clearly more extensive – it is found in most human cultures, so it exceeds tone languages in predominance – and it has been more comprehensively studied. Returning to IDS, then, we might argue with Dissanayake’s assertion (Dissanayake, 2000) that it began in *Homo heidelbergensis* and early *Homo sapiens* as an exclusively mother-infant communication and then spread more widely within hominin cultures to form the basis of musicality. Morley argues that such “social-emotive vocalization” – essentially HmMMM –

was a form of communication that came to be used *throughout* the social group at a much earlier time, without preference, both adult-adult and infant-adult, but is now perpetuated, in this predominantly non-lexical form, in adult-infant interactions and the prosodic content of adult speech. Furthermore, the shared prosodic pitch- and tempo-related properties of emotional vocalization (I[infant]D[irected] and A[dult]D[irected]) and music are not borrowed from one to the other, in either direction, but are, and always have been, a shared fundamental component of both (Morley, 2012, p. 127; his emphasis).

For Morley, as for Mithen, social-emotive vocalization originated towards the beginning of the *Homo* genus and not, with *Homo sapiens*, towards the end. Moreover, in broad alignment with Mithen's position, he argues that it "might gradually have evolved into music, as Dissanayake suggests, or at least provided shared foundations, but it could also have been the basis for language amongst all of a population" (Morley, 2012, p. 127).

### **3. Language and Cognition: Cognitivism *versus* Communicativism**

Having outlined how Hmmmmm might have become articulated into discrete segments, and how any segments which became freighted with meaning might have gone on to constitute language, I now consider certain issues in the philosophy of language which have a bearing upon later stages of this hypothesized process. Because one selection pressure driving the bifurcation of Hmmmmm into language and music was the need to communicate thoughts and desires, it follows that language is associated in some way with the thoughts it evolved to help communicate. I attempt therefore to deal now with the thorny question of the relationship between language, thought and consciousness, taking ideas of Peter Carruthers and integrating them with precepts from memetics. Adopting certain conclusions of Carruthers' will allow me, in Section 5, to argue for stronger functional similarities between the syntactic and semantic dimensions in language and music than have hitherto been acknowledged.

#### **3.1. Two Conceptions of Language in/as Thought**

Considerable debate surrounds the issue of how language and cognition relate to each other. Is language the mechanism for thought, the medium through which it is (exclusively) conducted (the so-called "cognitive conception" of language); is it simply a vehicle for, or translation of, thoughts conducted more fundamentally, in some kind of brain-language or "mentalese" (the so-called "communicative conception" of language); or does it occupy some intermediate position between these extremes (Carruthers, 2002, p. 657)? The cognitive conception of language is associated with the "relativism and radical empiricism" of Benjamin Whorf's (Whorf, 1956) view of language – "the Standard Social Science Model", in Steven Pinker's somewhat dismissive opinion (Carruthers, 2002, pp. 661, 664). By contrast, the communicative conception of language is generally more strongly advocated by cognitive scientists and evolutionary psychologists.

In part, the distinction devolves to one of nurture (cognitivism) *versus* nature (communicativism). For "cognitivists", such as Daniel Dennett (Dennett, 1995), the mind exists because the *tabula rasa* of the new-born child is shaped (bottom-up, inductively, *a posteriori*) by the nurtural power of language (indeed, in Dennett's view, by the power of memes themselves). For "communicativists", such as Steven Pinker (Pinker, 1997), much of the mind is naturally and innately pre-formed (top-down, deductively, *a priori*) at birth by natural selection, so memes, if they are implicated at all in cognition, do not do all the heavy lifting; rather, they act merely as epiphenomena of more fundamental processes. Seen in these terms, cognitivism intersects partly with "constructionist" approaches to language, which assert that "[g]rammar does not involve any [innate] transformational or derivational component"; rather, "learned [memetic] pairings of form [sound pattern] and function [meaning/concept]" constitute structures "in a network in which nodes are related by inheritance links" and in which "[s]emantics is associated directly with surface form (Goldberg, 2013, p. 15, see also 2003; Boas & Sag, 2012).

There is clearly no consensus on this particular nature-nurture question, despite the two positions

not being mutually exclusive; and responses to the issues involved tend, as suggested, to be split along disciplinary lines. A fuller understanding certainly requires an interdisciplinary integration of neuroscience, psychology and philosophy. The argument advanced in Carruthers (2002; see also the open peer commentaries (pp. 674–705) and Carruthers’ response (pp. 705–718)) is one of the most convincing attempts to unpick the issues involved, and his preferred analysis will be taken as the basis for what follows because of its ready accordance with the memetic interpretation advanced here. Carruthers, a moderate cognitivist, essentially attempts to chart a *via media* between cognitivist claims of different strengths, ranging from weak (language is necessary for at least some kinds of thought) to strong (language is essential for all types of thought) and, by doing so, implicitly considers the communicativist inversion of this continuum.

### **3.2. Thought, Modularity and Language**

Carruthers starts from the position that while “*some* thoughts are carried by sentences (namely, non-domain-specific thoughts which are carried by sentences of natural language), others [i.e., domain-specific thoughts] might be carried [non-linguistically] by mental models or mental images of various kinds” (Carruthers, 2002, p. 658; his emphasis). Carruthers’ hypothesis is that

[all] non-domain-specific [conscious and unconscious] thinking operates by accessing and manipulating the representations of the language faculty. More specifically, the claim is that [all] non-domain-specific [conscious and unconscious] thoughts implicate representations in what Chomsky ... calls ‘logical form’ (LF). Where these representations are *only* in LF, the thoughts in question will be non-conscious ones. But where the LF representation is used to generate a full-blown phonological representation (an imagined sentence), the thought will generally be conscious (Carruthers, 2002, pp. 658, 666; his emphasis).

To accept this, one has to endorse a modular view of mental structure similar to (but not necessarily in complete accordance with) the views expressed in (Fodor, 1983). In this account, “besides a variety of input and output modules (including, e.g., early vision, face-recognition, and language), the mind also contains a number of innately channeled conceptual modules, designed to process conceptual information concerning particular domains” (Carruthers, 2002, p. 663). These modules, for which strong selection pressures existed in early hominins, “include a naïve physics system ... a naïve psychology or ‘mind-reading’ system ... a folk-biology system ... an intuitive number system ... a geometrical system for reorienting and navigating in unusual environments ... and a system for processing and keeping track of social contracts” (Carruthers, 2002, p. 663).

By LF is understood here the unconscious mentalese structures underpinning and motivating the various connections possible between the components of natural language, in particular the relationships between verbs and the other sentence-elements required to combine with verbs in order to make the sentence grammatical, which some grammarians discuss under the rubric of “valency” (Durrell, Kohl, & Loftus, 2002, Chapter 8). This issue is considered further in Section 5.1.

As Carruthers argues, a LF, that is, “a non-conscious tokening of a natural language sentence would be ... a representation stripped of all imagistic-phonological features, but still consisting of natural language lexical items and syntactic structures” (Carruthers, 2002, p. 666). Such “imagistic-phonological” features would appear to equate to the language memes (lexemes) associated with a given LF. By “language memes” I refer to the imagined (internally heard) or spoken (physically produced) sound patterns of one or more words. This sense is broadly analogous to Saussure’s notion of “sound image” (see Section 5.2).

While domain-specific thought operates independently of language (using mental models or images), non-domain-specific thought, in being *tokened* (Carruthers, 2002, p. 660) by language, draws upon language’s syntactic structure – mediated by the underlying Chomskyan LF – to *constitute* it, not merely to *express* it (Carruthers, 2002, p. 664). Essentially, LF impels the generative/transformational aspect of language (Chomsky, 1965; Lerdahl & Jackendoff, 1983), whereby a finite set of recursive and hierarchical syntactic structures can underpin an infinity of content-specific utterances. In particular, Carruthers suggests that “distinct domain-specific sentences might be combined into a

single domain-general one” by means of “multiple embedding of adjectives and phrases” (Carruthers, 2002, p. 669), giving as an example “THE TOY IS IN THE CORNER WITH A LONG WALL ON THE LEFT AND A SHORT WALL ON THE RIGHT”, produced initially in mentalese as a mental model or image by a geometrical module; and “THE TOY IS BY THE BLUE WALL”, similarly produced by an “object property” module dealing among other things with colour. These become (unconsciously) integrated by LF as the basis for the domain-general, and potentially (consciously) lexemically manifested, “*the toy is in the corner with a long wall on the left and a short blue wall on the right*” (Carruthers, 2002, p. 669).<sup>6</sup>

Figure 1 (a visualization and extension of certain aspects of (Carruthers, 2002)) hypothesizes how the various input and output systems, and their associated modules, might be organized and how they might interact.

### Figure 1: Thought, Modularity and Language

The domain-specific modules – such as (naïve) physics, (folk) biology and (naïve) psychology, the latter termed here “ToM” (Theory of Mind) – are shown in the intermediate (middleground) layer.<sup>7</sup> While these and other modules are represented here as discrete “silos”, they are presumably highly interconnected in neurobiological reality. Moreover, while conceived in terms of input-output connections, modules also store information and so involve memory, of varying degrees of volatility. This memory is hypothesized to be encoded in the brain in accordance with the precepts of the “hexagonal cloning theory”, discussed in Section 4.

The domain-specific modules receive perceptual-sensory input processed by the hearing and vision centres (and also those for taste, touch and smell), shown in the background layer; and they can also backproject to these sensory inputs, as in situations where visual imagination is used to recreate or generate images and patterns (Carruthers, 2002, pp. 658, 666, 670). For clarity, not all linkages from sensory input to the domain-specific modules are shown in Figure 1. The language module, shown in the foreground layer, consists of comprehension and production sub-modules/systems and receives inputs from, and sends outputs to, the domain-specific modules. As Carruthers argues,

[The] production sub-system must be capable of receiving outputs from the [domain-specific] conceptual modules in order to transform their creations into speech. And its comprehension sub-system must be capable of transforming heard speech into a format suitable for processing by those same [domain-specific] conceptual modules. Now when LF representations built by the production sub-system are used to generate a phonological representation, in “inner speech”, that representation will be consumed by the comprehension sub-system, and made available to central [domain-specific] systems [, including the ToM] module.... perceptual and imagistic states get to be phenomenally conscious by virtue of their availability to the higher-order thoughts generated by the theory of mind system.... this is why inner speech of this sort is conscious: It is because it is available to higher-order [ToM] thought (Carruthers, 2002, p. 666).

In Figure 1, the *production* sub-system (“P”, and the associated blue arrows) is shown receiving outputs of the Number and Geometry modules after the receipt of some visual stimulus (purple arrows).<sup>8</sup> These mentalese inputs are synthesized into a LF which potentially serves as the foundation and cue for a lexeme – in this case, perhaps one articulating some notion of the quantity of a certain environmental shape or regularity. Whether verbalized or not (the former indicated by the arrow to “produced speech”), the production sub-system may generate a phonological representation in “inner speech” (the lexeme sounding internally, perhaps by recruiting auditory-system neurons). Over time, and as a result of enculturation, evolutionarily stable associations (co-adaptations) between certain LFs and certain lexemes – in a kind of “lock-and-key” process – constitute language learning. This phonological representation is “consumed” by the comprehension sub-system (“C”, and the associated green arrows). Its availability to higher-order thought via the ToM module (indicated by the arrow from the comprehension sub-system to the ToM module) renders it conscious, even though (as Carruthers’ remarks might be taken to imply) consciousness (and therefore language) is not necessary for comprehension.<sup>9</sup> This “zone of consciousness” is approximated by the dotted circle in Figure 1.

As far as language *reception* is concerned, perceived speech (red arrows, initially from “Auditory

Input (Heard Speech)”) is directed towards the Comprehension sub-system via the hearing centre and cognized by means of “deconstruction” of its inferred LF into “mental models or mental images of various kinds” and by reference to the relevant domain-specific modules necessary to understand it. In the case of Figure 1, these are Biology and Number – appropriate for a sentence articulating some notion of the quantity of a particular animal or fruit.

Having explained how underlying LF mentalese may be associated with an “imagistic-phonological” lexeme, I argue in Section 5.3 for a musical equivalent to this process: an association between LF mentalese and similarly “phonological” – but perhaps less overtly “imagistic” – musemes.

### **3.3. Memetics, Cognitivism and Communicativism**

Where does memetics fit into the cognitivism-communicativism debate? Its adherents would appear naturally to gravitate towards the cognitivist perspective, given their belief that education and enculturation load the brain with what might be termed “verbal-conceptual memes” (of which more later) and, in a computational metaphor, impart to it the software to augment the evolution-derived capacity of the underlying biological hardware. This is essentially Dennett’s view – the brain as a “Joycean machine” illuminated by memes (Dennett, 1993, p. 214; see also Rice, 1997). As he expresses it,

[h]uman consciousness is itself a huge complex of memes (or more exactly, meme-effects in brains) that can best be understood as the operation of a [serial] virtual machine implemented in the parallel architecture of a brain that was not designed for any such activities. The powers of this virtual machine vastly enhance the underlying powers of the organic hardware on which it runs, but at the same time many of its most curious features, and especially its limitations, can be explained as the byproducts of the kludges [*ad hoc* software bug repairs] that make possible this curious but effective reuse of an existing organ for novel purposes (Dennett, 1993, p. 210; his emphases).

Thus while not denying that some kinds of thought are possible without language (Carruthers, 2002, p. 661), Dennett is adopting the strong cognitivist perspective that a whole new vista was opened up by the lexemes which in his view constitute (as opposed to reflect) the majority of human thought. For him, these lexemes are the raw materials whose manipulation constitutes (as opposed to merely expresses) the substantial majority of thought.

But – and *contra* Dennett – adopting some flavour of communicativism does not necessarily mean abandoning a memetic view of culture, thought or consciousness: it simply means accepting that some types of mental content are not *directly* amenable to the kind of imitative lexemic transmission upon which memetics is predicated, and therefore that direct cultural transmission is only part of the picture. The remaining elements, reasonably enough, are to be sought in those areas of mental functioning shaped most strongly by biological evolution – nature as opposed to nurture. Of course, a co-evolutionary viewpoint maintains that cultural evolution exerts significant pressure upon biological evolution.

From this standpoint – and assuming a distinction is maintained between domain-specific mentalese (representing, as noted, “mental models or mental images of various kinds” (Carruthers, 2002, p. 658)) and domain-general, LF-underpinned language – it is the latter which not only facilitates a synthesis of the former but which also offers a means of (memetically) transmitting the integrated information content between individuals by means of the tokening of LF structures by lexemes and their subsequent dissemination. So not only do we have i) the traditional brain-to-brain *transmission* of classical memetics, but also ii) an intra-brain *translation* process between those (unconscious) mental structures which encode domain-specific mentalese and integrated LF mentalese – each a *mnemon*, or item of (initially unreplicated) brain-stored information (Lynch, 1998) – and those (conscious) structures which encode domain-general lexemes. This distinction is represented in part by the coloured meme-symbols in Figure 1.

So even a strong communicativist position is not incompatible with memetics, because even though thought is not taken in this view to be *implemented* by language, language is still the medium

by which certain (integrated) thoughts are *transmitted* between individual brains (with concomitant reconstitution back to mentalese in the receiving brain). Thus, for memetics, the differences between cognitivism and communicativism, strong or weak, tend to devolve to what exactly is translated into, and reconstituted from, language. Moreover, as I argue in Sections 4 and 5, the neurobiological structures encoding the various types of mental content are not necessarily different in kind, only in the type of information (mentalese LF, lexemes, musemes) they encode.

While memes are perhaps most readily supported in the realm of language on the grounds that linguistic utterances may be relatively easily imitated (for Kirby, “language appears to have adapted simply through the process of iterated learning in such a way as to become more learnable” (Kirby, 2013, p. 129)), certain domain-specific thoughts might still be memetically transmitted. In the case of Carruthers’ “long wall on the left and a short wall on the right” example (Section 3.2), a mnemonic (in mentalese) encoding this information (in the sense of highlighting it as a specific conceptual entity) might be transmitted from one individual to another by means of gestures and facial expressions in the context of sensory stimulus (a visual input of the long and short walls), potentially bypassing any linguistic formulation entirely. If the stimulus-wall is coloured, and if this information is transferred, then a domain-*general* meme will have been transmitted. In this way, the mnemonic becomes a meme by virtue of its reconstitution in another brain by means, if not strictly of imitation, certainly of a form of social learning. This is broadly analogous to the “stimulus enhancement” found, for example, in tits who observe others of their species pecking at milk-bottle tops (this behaviour being reinforced by the acquired cream) and, their attention thus directed to the bottle (the stimulus), themselves repeat the behaviour (Blackmore, 1999, p. 48).

## **4. Calvinian Mechanisms for Information Encoding**

Understanding language, music and thought requires not just interdisciplinary interaction between biology and philosophy but also a discussion situated at a variety of explanatory levels. Here we turn to the level of the neuron in order to account for how the kinds of musical and linguistic information discussed in Sections 2 and 3 might be encoded in the brain. This affords us a means by which a synthesis of these issues can be attempted (Section 5).

### **4.1. The Hexagonal Cloning Theory**

Is there a known mechanism of neural information encoding which might allow us to mediate between memes in language and those in music, and to accommodate Carruthers’ view of language as a medium for non-domain specific thought? One candidate is a family of related theories which stem ultimately from Donald Hebb’s work in the 1940s on the columnar organization of neurons. Hebb argued, and subsequent work has confirmed, that rather than being organized randomly within the cerebral cortex, certain cells – the pyramidal neurons – are organized into discrete columns, each of which is implicated in the encoding and representation of an element of perception or cognition (Hebb, 1949). Subsequent research supported this hypothesis (Mountcastle, 1978). Many investigators have observed empirically that such cells tend to form co-resonating arrays in the geometrically optimal form of the triangle, their interdigitation allowing several attributes of a percept or thought, each associated with a specific triangular array, to be represented (Leng, Wright, & Shaw, 1990; Leng & Shaw, 1991).

Calvin has extended these theories, arguing that such triangular arrays themselves form hexagonal plaques on the surface of neocortex (Calvin, 1996; see also Jan, 2011). Moreover, to this geometric extension he posits the element of Darwinian competition between rival hexagons, for the optimal encoding of a multi-component percept or thought (Fernando, Szathmáry, & Husbands, 2012). While the argument of this article is not contingent upon there being a specific topography of neuronal structures (it requires only that discrete phenomena in the world are encoded discretely in the brain), subsequent work on spatial location encoding in the entorhinal cortex (Fuhs & Touretzky, 2006; Burak & Fiete, 2009; Doeller, Barry, & Burgess, 2010; Mhatre, Gorchetnikov, & Grossberg, 2012;

Stensola et al., 2012) has supported Calvin's ideas. Indeed, this research, accounts of the tonotopic organization of the auditory cortex (Zatorre, 2003, p. 233), and of the phototopic organization of the visual cortex (Braitenberg & Braitenberg, 1979; Reichl et al., 2012a, 2012b) not only suggest deep similarities between brain representations of a variety of sensory inputs they also indicate that, for all the astonishing complexity of neuronal interconnections, a triangular-hexagonal disposition of cortical columns is a recurrent structural-topographical configuration.

Because incoming perceptual information is often pre-segmented into discrete units by gestalt processes operating at "lower" levels of the perceptual input system (represented by the background level of Figure 1), the auditory data encoded by cortical hexagons may constitute (but is not necessarily limited to) memes and musemes (Jan, 2011, sec. 4.1.1). If so, it is useful to make a distinction between this brain-encoded form of a meme which, after "genotype", one might term the "memotype", and certain extrasomatic, physical artefacts and behaviours to which the memotype is capable of giving rise and which, after "phenotype", one might term the "phemotype" (Jan, 2007, p. 30, Table 2.1).

## **4.2. Hemispheric Localization of Music and Language**

Significant progress has been made over the last decade in understanding the localization of musical and linguistic function in the brain and the resulting knowledge aligns well with the evolutionary account of HmMMMM bifurcation offered by Mithen. To summarize (and inevitably oversimplify) a complex picture, structures in the right hemisphere appear to dominate the processing and generation of contour, tonality and timbre of both melody and speech; whereas structures in the left hemisphere appear to dominate the processing and generation of syntactic organization and semantic content in language, together with rhythmic structure in both language and music (Morley, 2012, p. 118; but see Patel, 2008, pp. 73–6).

Thus, structures in both hemispheres are involved in the production and processing of both music and language; some of the fundamental elements of music and language production and perception are shared ... and some have subsequently become specialized. Musical functions are a whole are less clearly lateralized than language function, but tasks relating to pitch and pitch discrimination do seem to be right-hemisphere dominated. Linguistic functions seem to be most detrimentally affected by left hemisphere lesions; most musical functions seem to be impaired in some respect by damage to either hemisphere (Morley, 2012, p. 118).

This view is broadly supported by Brown, Martinez and Parsons (2006). On the basis of PET scans of subjects engaged in sentence and melody generation/completion tasks, they argue for i) shared (and therefore co-localized) neural processing of certain music and language features; ii) for parallel processing and partial overlap (and therefore some co-localization) in brain systems for certain other features of music and language; and iii) for distinct processing (and therefore separation) in brain substrates for yet other music and language elements (Brown et al., 2006, p. 2798, Fig. 5). They conclude that

[w]hereas music and language may share resources for audition and vocalization, phonological generativity is seen as the major point of cognitive parallelism between them, in which parallel cognitive operations related to combinatorial phrase generation occur on divergent semantic units (Brown et al., 2006, p. 2801).

From an evolutionary perspective, it would appear that the neural substrates for HmMMMM were, as an essentially melodic-rhythmic phenomenon, bi-lateral; and as segmentation and compositionality evolved, the substrates responsible for language-primary elements (syntax and semantics) were increasingly focused in the left hemisphere.<sup>10</sup> As Martin and Perry argue, "speech may have evolved from an already-complex system for the voluntary control of [HmMMMM/social-emotive] vocalization. Their divergences suggest that the later evolving aspects of these two uniquely human abilities are essentially hemispheric specialisations" (in Morley, 2012, p. 119).<sup>11</sup>

Understanding this specialisation in the light of the HCT, it might be hypothesized that, over the course of hominin evolution, right-hemisphere hexagonal plaques representing increasingly discrete

(FOXP2-segmented?) sonic units (protemes) were yoked (by means of connections to be discussed in Section 5.1) to left-hemisphere hexagonal plaques regulating their syntactic interrelationship and semantic content – these perhaps even implementing a proto-LF – in order to engender the compositional lexemes of language.

## **5. Towards a Memetic-Carrutherian-Calvinian Synthesis for Language and Music**

Having discussed the possible mechanisms underpinning the bifurcation of Hmmmmm into language and music (Section 2), Carruthers' hypothesis on the relationships between language and thought (Section 3), and the neural substrates which might underpin these various phenomena in different parts of the brain (Section 4), I now turn to discuss how these various perspectives might be synthesized. My aim is to explore certain functional parallels between language and music. Specifically, by considering syntactic and semiotic relationships between these domains, analogies – arguably, homologies – between them might be more fully understood and Mithen's hypothesis on their evolutionary relationships to the parent Hmmmmm might therefore be further substantiated.

### **5.1. Implementation of Linguistic Syntax in the Light of the HCT**

Carruthers' suggestion (Section 3.2) that “distinct domain-specific sentences might be combined into a single domain-general one” by means of “multiple embedding of adjectives and phrases” (Carruthers, 2002, p. 669) – a means for the implementation of his central hypothesis – has a ready mechanism in the HCT. Calvin suggests that hexagons encoding certain kinds of mental data in one part of cortex are connected to others encoding different kinds of data in a different region. Moreover, he argues, drawing on ideas of Antonio Damasio's, that “there are specialized places in the cortex, called ‘convergence zones for associative memories’ – or ‘association cortex’ – where [representations in] different modalities come together” (Calvin, 1996, pp. 129–130). Calvin speaks of “hashing” or abstracting the attributes of a “distributed [domain-specific] ‘data base’” in order to create a “centrally located [domain-general] representation”, the mechanism for which appears to be index-hexagonal overlapping/interdigitation in association cortex (Calvin, 1996, pp. 17, 207, 135).

The connections between domain-specific hexagonal codes (a sub-committee, to adapt one of Calvin's metaphors (Calvin, 1996, p. 45)) and the fully “associated” domain-general LF code (a master committee) are achieved by certain types of “corticocortical projections” which go beyond the localized connectivity responsible for supporting triangular/hexagonal arrays and which involve links which “can go long distances, as from one hemisphere to another ..., though most only make a U-shaped passage through the white matter of one gyrus and then terminate in a nonadjacent patch of cortex that's only a few centimeters away” (Calvin, 1996, p. 131). Because such links are able to reconstitute the hexagonal plating of one area of cortex in another, Calvin terms them a “*faux fax*” and, writing in the mid-1990s, likens them to hyperlinks in the then nascent internet (Calvin, 1996, pp. 125, 131).

The following is an overview of how certain key aspects of language syntax are implemented by the HCT and *faux-fax* linkages.

1. The adjectival modification of a noun may be accounted for by “simple borderline superposition of hexagons” (Calvin, 1996, p. 193). Beyond a certain point (several adjectives and, perhaps, prepositions), however, superposition runs the risk of creating an unspecific – Bickertonian-protolinguistic (Calvin, 1996, p. 193) – mix of words, the solution to the potential chaos of which is hierarchical recursive embedding (see point 4 below).
2. The binding of a pronoun to its referent may be accomplished by a *faux-fax* link which connects the representations of these two words, even if they are in different sentences (Calvin, 1996, p. 194).

3. The long-range dependencies of wh- questions are similarly implemented (Calvin, 1996, p. 194). The assumption for both point 2 and point 3 is that the *faux-fax* is bidirectional. Using the metaphor of a choir, Calvin argues that “[b]ack projections ... can use the same code, and so immediately contribute to maintaining a chorus above a critical size.... A backprojected spatiotemporal pattern might not need to be fully featured, nor fully synchronized, to help out with the peripheral site’s chorus” (Calvin, 1996, p. 194).
4. Recursive embedding – which is “at the very top of [linguists’] Universal Grammar wish list” (Calvin, 1996, p. 194) – is implemented by *faux-fax* links which allow higher-level concepts to connect representations of subsidiary parts of a sentence intelligibly. According to Calvin, “if either subchorus [a discrete clause] falters, the top-level one [the integrity and sense of the sentence as a whole] stumbles” (Calvin, 1996, p. 194). Calvin gives the example of the sentence “I think I saw him leave to go home” (computationally/hierarchically, X://I think/I saw him/leave/to go/home), wherein the Darwinian success of the hexagonal colonies representing the top-level *think* verb is dependent upon the survival of the *saw* and *leave* verb colonies connected to it via *faux-fax* links. In a process of “stratified stability”, “[i]f the *leave* link stumbles, the *saw* hexagons might not compete very effectively and so the top level [*think*] dangles” (Calvin, 1996, p. 195). For this system to work, “[e]ach verb has a characteristic set of links: some required, some optional, some prohibited” (Calvin, 1996, p. 195) – termed valency in Section 3.1.

Such connections and their associated hierarchic relationships would appear to be key to the nature of LF. Moreover, the various references to specific parts of speech here apply primarily to their LF representations and only secondarily to the associated (tokening) lexemes.

To summarize, the HCT (and with it *faux-fax* linkage and the Darwinian competition between cortical hexagons) is a candidate mechanism for Carruthers’ central hypothesis of language as the medium for domain-general thought. This is because (building upon the discussion of Section 4.2) it affords a means by which hexagons encoding domain-specific representations of “mental models or mental images” (in various regions of the brain) can be interconnected to (left-brain-situated?) domain-general/LF conglomerations. These LF structures can then be similarly associated with those (right-brain-situated?) hexagons encoding the co-adapted lexemes.

## **5.2. Semantic Homologies Between Language and Music**

One might extend and support the discussion in Section 5.1 by considering how musemes might also bear semantic content by virtue of mechanisms analogous to those linking linguistic LF structures – which integrate domain-specific meanings – to lexemes.<sup>12</sup> In this sense music is seen as acting as a kind of degraded language, retaining some of the semantic capacity of Hmmmmm by virtue of its ability, like the sound patterns of its antecedent, to become associated (sometimes arbitrarily, sometimes not) with extra-musical concepts, but lacking the kind of rich, semantically implicative syntax of language. Of course music has its own rich syntax, but this is, to use a distinction of Agawu’s (after Roman Jakobson), generally more “introversive” than “extroversive” – so whereas the inversion of words in a sentence might have global semantic effects, a comparable inversion in music might only be understood to perturb the local syntax (Agawu, 1991, p. 23). I consider this issue further in Section 5.3, arguing, nevertheless, that the LF structures which lexemes token might have an analogue/parallel in music, their neural substrates being partially interconnected.

To help focus the discussion, I concentrate here primarily on the “topics” of late-eighteenth century music (Agawu, 1991; Ratner, 1991; Allanbrook, 1992; Caplin, 2005; Monelle, 2006), which, in Meyer’s terms, are broadly understood and widely held “connotations” afforded by musical patterns (Meyer, 1956, p. 258). Topics are formed by educated listeners from the historically contingent, indexical connections between certain musical patterns and specific extra-musical ideas. The former include dance-associated rhythmic sequences (“types” (Ratner, 1980, p. 9)), together with more in-

tangible associations of pitch and texture (“styles”); the latter include generic notions of social hierarchy and specific concepts and images. The mechanisms which afford semantic content to topics seem applicable to more private associations, such as those individual composers and listeners might form between particular pieces of music and certain extra-musical ideas, and so they may be generalizable beyond the frame of reference considered here.

One means of mediating between music and language in this respect is through classical semiology, specifically its association of a *signifier* with a *signified*. As Saussure argued in his celebrated definition,

[t]he linguistic sign unites not a thing and a name, but a concept [the signified] and a sound-image [the signifier]. The latter is not the material sound – a purely physical thing – but the psychological imprint of the sound, the impression that it makes on our senses: the sound-image is sensory, and if I happen to call it “material”, it is only in that sense, and by way of opposing it to the other term of the association, the *concept*, which is generally more abstract (in Nattiez, 1990, p. 3).

Mapping this onto the two conceptions of language and thought of Section 3.1, the following assertions might be made:

1. In the communicativist view of language, which aligns elegantly with Saussure’s definition, the “concept” is a domain-general, LF-implemented (unconscious) thought, whereas the “sound image” is one or more internally-heard (conscious) lexemes (and, it is argued, *musemes*).
2. In a cognitivist interpretation, which arguably aligns less well with Saussure’s definition, the “concept” (broadly speaking the function, in constructionist terms) would be regarded as existing purely (and simultaneously) in the shape of one or more (presumably) unconsciously active and consciously internalized lexemes (the constructionist form), and not as a LF.

Figure 2<sup>13</sup> generalizes the topical association between a *museme*  $m$  and a lexeme  $l$  (or with a “lexemeplex” or complex of lexemes). By this I mean that  $m$  is functioning in a broadly equivalent manner to  $l$ , in that both are internal/external sound-sequences which have the capacity to token LF-underpinned semantic associations. How one conceives the detailed operation of this process is nevertheless dependent upon whether one adopts a cognitivist or a communicativist standpoint; as noted earlier, the latter is adopted here:

1. From a cognitivist viewpoint, because most or all thought is understood to be conducted by means of the manipulation of language, any semantic content which might be possessed by  $m$  is wholly parasitic upon language, as the more fundamental medium.
2. From a communicativist viewpoint,  $m$ ’s semantic content may i) draw *indirectly* – i.e., via or mediated by language – upon the *semantic* elements of LF *mentalese*; but it may also ii) draw *directly* – i.e., unmediated by language – upon the semantic elements of LF *mentalese*.

As will be argued in Section 5.3, music may also to some extent draw directly upon the *syntactic* element of LF *mentalese* for its recursive-hierarchical structuring, in a manner which parallels language’s recursive-hierarchical structuring by LF *mentalese*.

**Figure 2: The Memetic-Semiotic Nexus of an  $m-l$  Music-Language M(us)emeplex**

Figure 2 is organized according to three different dimensions. As will be evident as the discussion progresses, these dimensions relate in various ways to the hemispheric localization of music’s and language’s neural substrates, discussed in Section 4.2. One of these dimensions is semiotic, in that it attempts to represent three distinct semiotic levels, termed “Level One”, “Level Two” and “Level Three”. Another dimension represents the memotype-phenotype (somatic-extrasomatic) distinction

outlined in Section 4.1, whereby a (bold-type) formulation such as “*m*” refers to the memotypic form of a museme *m* and where “ $\boxed{m}$ ” refers to its phemotypic expression. Note that the memotypic level is in principle conscious and is to be distinguished from the unconscious mentalese/LF structures with which it is associated and which it tokens. The third dimension makes a distinction between the two evolutionary outcomes of Hmmmmm, music and language.<sup>14</sup>

In Figure 2 part i a, columns 1 and 3, and at the lowest level of referring,  $\boxed{m}$  – the physical sonority through which *m*, via the intercession of voices (or musical instruments), impinges upon us most directly – is represented, in a “horizontal” memetic-semiotic relationship, as the phemotypic (coded-for) meme-product of the memotypic (coding-for) *m*. Thus,  $\boxed{m}$  acts as a (somewhat abstract) signifier for *m*.  $m \leftrightarrow \boxed{m}$  is often associated with a “grapheme”  $Gm \leftrightarrow G\boxed{m}$ , which partly governs the arguably superficial (from Carruthers’ point of view) matter of notating *m* and which, while not essential for its existence, is nevertheless (in the case of literate cultures) often significant for its transmission. The same principle is true, of course, in the case of lexemes.

By analogy with  $m \leftrightarrow \boxed{m}$ , columns 2 and 4 of Figure 2 i a illustrate analogous relationships for the lexeme *l*, which codes for the spoken expression  $\boxed{l}$ . Paralleling  $Gm \leftrightarrow G\boxed{m}$ ,  $Gl$  is a grapheme coding for the written expression  $G\boxed{l}$ . As with the music related memes, the phemotypic forms  $\boxed{l}$  and  $G\boxed{l}$  act as signifiers (again somewhat abstractly) for the associated memotypic signified forms *l* and *Gl* respectively.

As represented in Figure 2 i b, columns 1 and 3, and at an intermediate level of referring,  $Gm$  also exists, now as a *signifier*, in “vertical” semiotic co-adaptation with *m*, even though it is essentially independent of it (their relationship is “arbitrary” (Nattiez, 1990, p. 4)).  $\boxed{m}$  is similarly associated, as signified, with the corresponding phemotypic signifier meme,  $G\boxed{m}$ .

Analogously, *l* and *Gl* function as signifiers of the signified language “interpretant-lexemeplex” *Il*. By this is meant the wider network of cognate lexemes which provide the context for *l* and which anchor it in a broader web of signification.<sup>15</sup> The components of *Il* ultimately devolve, in a communicativist view, to the “back-end” LF-integrated “mental models and images” for which *l* (and *Il*) are the “front end”. In this sense, *Il* is the essence of the “conscious propositional thought” (Carruthers, 2002, p. 664) tokened by *l*. As with the *m*-related memes,  $\boxed{l}$  and  $G\boxed{l}$  function as signifiers of the signified  $Il$ .<sup>16</sup>

As represented in part ii of Figure 2, and at the highest level of referring, the “diagonal” association between  $m \leftrightarrow \boxed{m}$ , as signifier, and  $Il \leftrightarrow \boxed{Il}$ , as signified, forms a *m-l* m(us)emeplex, one either confined to a particular individual or shared more widely (topically) within a cultural community. In such associations, the presence of the musical element triggers/cues the verbal in consciousness (or vice versa). In this sense, level-three semiosis corresponds not only to scenario 2 i in the (second) list earlier in this subsection, but also potentially to scenario 2 ii – that is, the linking of musemes (directly) to the *semantic* elements of LF mentalese, displacing (or supplementing, in an intermediate state between scenarios 2 i and 2 ii) their normal lexemic token. Such “semantic elements” are the interconnected mentalese codes for nominal, adjectival, verbal, etc. functions – the “natural language lexical items and syntactic structures ... stripped of all imagistic-phonological features” (Carruthers, 2002, p. 666) – which constitute LF.

This might be particularly the case with musemes which, on account of their strong image-schematic/embodied properties (Cooke, 1968; Snyder, 2000, p. 108ff.), link primarily iconically with LF representations deriving from one or more of the domain-specific modules of Figure 1. Nevertheless, in the case of topics, indexical linkages might also arise, because many topics have real-world (albeit not always arbitrary) referents underpinning them, such as the emulations of horn and trumpet dotted rhythms which constitute the “military” style, or the bagpipe-like drones which define the “musette/pastorale” style (Ratner, 1980, pp. 18–19, 21). In such cases, a context in which the instrument (or the dance rhythm, in the case of Ratner’s rhythmic types) is used affords meaning to the topic.

The various cells in Figure 2 are connected by double-headed arrows, which represent the associations or linkages between phenomena in different dimensions and substrates by which understanding and meaning emerges. While the representation of patterns and their linkages on a two dimensional

page is useful to foster clarity of exposition and discussion, it also appears the case that this mirrors, to some extent, real functional and structural localization and interconnection in the brain. As intra-brain linkages, all the vertical and diagonal connections linking columns 1 and 2 of Figure 2 (shown as red arrows) can potentially be accounted for by the HCT (Section 4). Naturally, the horizontal connections from columns 1 to 3 and 2 to 4, and the vertical and diagonal connection between columns 3 and 4 (shown as blue arrows) cannot be accounted for in this way, because they are not intra-brain linkages but rather somatic-extrasomatic (inter-brain) associations. In the case of columns 1 and 2, however, the red double-headed arrows are the graphical equivalent of the *faux-fax* links which Calvin argues connect representations in one region or functional domain of the brain with those in another.

### 5.3. Implementation of Musical Syntax in the Light of the HCT

If the communicativist view of language is one of (left-brain) LF translated into and reflected by imagined and spoken (right-brain) lexemes, could (introversive/syntactic) musical “thought” also be conducted in a form of mentalese – a (left-brain) LF grammar of music – before association with the (right-brain) musemes which give rise to imagined and heard music? This question is an extension of point 2 ii (the potential for music to “draw *directly* – unmediated by language – upon the semantic element of LF mentalese”) in the (second) list given in Section 5.2, whereby not (just) the *semantic* but also the *syntactic* element is drawn upon. As noted, these two dimensions are closely interconnected in language, but in music they are more independent.

It seems the case that processes covered under point 4 of Section 5.1 might also account for the representation of grammatical-hierarchical structure in music (Lerdahl, 1992; Temperley, 2001). In the same way that “*faux-fax* link[s] ... [allow] higher-level concepts to connect representations of subsidiary parts of a sentence intelligibly” (to form a fully associated domain-general LF code), they might also connect subsidiary parts of a musical phrase together under some overarching “higher-level concept”, such as a framework harmonic progression, a “structural-melodic line” (Ratner, 1980, p. 89, Ex. 6–7), a Schenkerian *Zug* (Schenker, 1979, pp. 43–46), or some other schema (Leman, 1995; Gjerdingen, 2007; Byros, 2009).

Moreover, in the same way that the multiply embedded clausal structure of a sentence is presumably replicated at a recursively higher level across a number of sentences, the same may be true for music. Deliège’s notion of “cue abstraction” or Gjerdingen’s concept of “*Il filo*” (the “thread”, along which a discrete series of schemata are arranged) might be a candidate psychological manifestation of this neurobiological process (Deliège, 2000; Cambouropoulos, 2001; Gjerdingen, 2007, p. 369; see also Jan, 2010). In this sense, music’s syntax – which has been the subject of extensive speculation ranging from the rhetorical schemata of the seventeenth century (Bonds, 1991) to the Chomskyan speculations of the 1980s (Lerdahl & Jackendoff, 1983) – might, as suggested in Section 5.2, be partly dependent either upon some degree of interconnection between (linguistic) LF and a dedicated musical analogue, musical LF, perhaps proximately located in the brain.

While further research is needed (such investigations being to some extent contingent on ever finer resolution in neuroimaging technologies), there is some evidence for this, in that Brodmann’s areas 44 and 45 appear to implement a parallel “syntax/phonology interface area” subserving these functions in both domains (Brown et al., 2006, p. 2798, Fig. 5). Moreover, Patel proposes a “shared syntactic integration resource hypothesis” (SSIRH), which posits that language and music “have distinct and domain-specific [parallel] syntactic representations (e.g., chords vs. words), but that they share neural resources for activating and integrating these representations during syntactic processing” (Patel, 2008, p. 268). The mechanism for this activation/integration in music might thus involve the same kind of (*faux-fax*) connections between right-hemisphere music centres and left-hemisphere semantic-syntactic LF centres discussed in Section 4.2. This reinforces a view that, post-bifurcation, both domains retain significant structural and functional homologies, because it was evolutionarily inefficient for them wholly to separate in their input, representation, or output systems.

The argument for an LF grammar of music runs as follows, and requires three coordinated “ifs”.

If (i) music and language did share a common ancestor in the form of HmMMMM; if (ii) sonorously depleted but semantically rich language is a reflection of an underlying brain language (the communicativist claim); and if (iii) the latter attribute was present originally in HmMMMM, then sonorously rich but semantically depleted music could have retained some element of this communicativist attribute. In this way, both evolutionary descendants of HmMMMM might be reflections of an underlying LF mentalese.

The third “if” is perhaps the most problematic in that, in its archetypal form, HmMMMM (as discussed in Section 2) is a syntactically undeveloped form of communication, lacking the compositionality of grammatically developed language. As Carruthers argues, “it is natural language *syntax* which is crucially necessary for inter-modular integration” (Carruthers, 2002, p. 658; his emphasis). If his model is taken to hinge upon the underpinning and constitution of language by some form of mentalese-level, syntax-articulating LF, then perhaps the non-compositional HmMMMM does not in fact implement it, and the argument for any evolutionarily persisting communicativism in music therefore falls. But if some form of communicativism does not require a *fully* developed syntax – if, in other words, it allows various shades of syntax, including the “protosyntax” potentially underpinning later, more developed forms of HmMMMM (and, indeed, Bickertonian protolanguage) – then HmMMMM, and with it its evolutionary descendant music, might indeed be amenable to a communicativist interpretation.

The latter would appear to be the more likely scenario, because HmMMMM presumably evolved relatively smoothly and gradually into language and music over many millennia, and not by means of “saltationist” jerks. This accords with the general view in evolutionary theory that even a little bit of a good thing is preferable to none of it (Dawkins, 1991, p. 90). One piece of evidence in favour of such “shades of syntax” might be derived from the earlier discussion on segmentation (Section 2.1). Once the processes engendering segmentation had started to have their effect on HmMMMM, one would be in a transitional phase – one presumably lasting many hundreds of thousands of years – where attributes of both older HmMMMM and newer compositional language were simultaneously present, HmMMMM acting as a framework or scaffold for the newer form of communication before finally being supplanted by it. The argument advanced here is that this “post-HmMMMM” possessed just enough syntax (proto-LF; Section 4.2) to give rise both to compositional language, communicatively understood, and to music *evolving on the basis of an underlying communicativist dualism between its own form of LF mentalese and imagined (musemic) sound*.

If, on the basis of the above, the third “if” is held to be true, then both language and music implicate some kind of (parallel, partially overlapping) LF representation. In language, this may be described by Chomsky’s generative-transformational grammar. In the literature of music theory, there are numerous accounts of the grammatical basis of music, with one in particular, Lerdahl and Jackendoff’s Generative Theory of Tonal Music (GTTM) (Lerdahl & Jackendoff, 1983) being the most explicitly linguistic, although to some extent chronologically/stylistically circumscribed.

## **6. Conclusion: Music and ‘Inexpressible Longing’**

To summarize the main conclusions of this article, the following has been argued:

1. Music and language are effectively two sides of the same evolutionary coin. Once the neural substrates for auditory stream segmentation were in place, it was inevitable that the chunks of sonorous information arising from HmMMMM would be subject to Darwinian replication, variation and selection. Computer simulation of the process offers telling evidence of its likely validity.
2. The replicated sound patterns of language are arguably proxies of a more fundamental mental language, LF. This fosters the integration of concepts in different domains to form a multi-modal syntactic-semantic complex which, in conjunction with language, is not only amenable to consciousness but which confers significant evolutionary advantages upon individuals who possess this facility.

3. Lexemes and musemes may be encoded in the brain in broadly similar ways – by means of hexagonal encoding, cloning and Darwinian competition – and they are predominantly right-hemisphere localized. This constitutes further evidence for their common evolutionary origin in HmMMMM. The syntactic structures encoding LF appear to be predominantly left-hemisphere localized. *Faux-fax* links connect the two types of representation, allowing the cross-hemispheric tokening of LF by lexemes.
4. The mechanisms by which language acquires semantic content appear broadly replicated (albeit more loosely) in music, and might be understood in terms of multi-level semiotic process spanning different replicator domains (memotypic, phemotypic). Moreover, it may be the case that LF structures, or analogues of them (in the left or the right hemisphere) might also subserve music's syntactic structure.

Two points arising from this list might be developed briefly in conclusion. One potentially fruitful extension of point 1 would be to examine how the sonic (perhaps in conjunction with the syntactic and semantic) attributes of lexemes correlate with their replicative success. In the case of musemes, it seems clear that the perceptual-cognitive salience of a musical pattern correlates strongly with its replicative-evolutionary fortunes (Jan, 2007). The most salient and striking musical patterns – perhaps those with the most interesting melodic contours or tonal structure – are normally those which are replicated most, which go on to appear in numerous musical works, and which therefore play the largest role in shaping the profile of the wider musical dialect. In this sense their perceptual-cognitive salience, however it is measured, is an index of their likely statistical prevalence in a given museme pool and, ultimately, of their “selfishness” (Dawkins, 1989; Distin, 2005).

While lexemes replicate under tighter syntactic and semantic constraints than musemes (in the sense that their mutation rate is limited to a greater extent by the imperatives of communication), it appears likely that, as segmented sound-units, they warrant consideration in similar ways to musemes. As with the origin, florescence and senescence of musical styles, genres and systems of tonal organization (Jan, 2013, p. 152, Figure 1, n.d.), the notion of linguistic “speciation” – recognized by Franz Bopp (1791–1867) before that in nature (J. Miller & Van Loon, 2010, p. 100) and adopted by Darwin as a means of illustrating biological speciation (Darwin, 2008, p. 311) – might be understood as a system-level consequence of the operation of the evolutionary algorithm upon the relevant unit of selection, the lexeme. Indeed, Dawkins gives a small but telling example of this in the mispronunciation of the second line of the chorus of “Rule Britannia” as “Britannia, rule[s] the waves”. This, he argues, is the result of the greater salience of the sibilant ending of “rules” as against the original “rule”; and also the more grammatically comprehensible indicative mood of the “rules” version, as against the more nuanced imperative, or even subjunctive, implication of “rule” (Dawkins, 1989, p. 324).

On point 4, if the first part of this is true, then one might ask why music is not as semantically specific as language. One reason might be that what might be termed an evolutionary “wedge” effect came into play after the bifurcation of music and language. That is, after separation their evolutionary paths diverged ever more widely because of the need for compositional language, as the information-communicating successor to HmMMMM, to remain broadly coherent and specific to all members of a socio-linguistic group, and the concomitant relaxation of this constraint upon music once language had begun to bear this burden.<sup>17</sup> Put another way, the “Humboldtian” nature of language – its compositional recombination of a relatively small number of component elements into a near infinity of utterances – developed along more syntactically circumscribed lines than was the case in music (Merker, 2002).

Freed of its obligation to communicate specific conceptual/propositional thought, music was increasingly able to fulfil less tangible – but no less evolutionarily important – roles, particularly the fostering of group cohesion through (holistic and multimodal) communal physicality and pleasure, still alive today in the throbbing beats of clubs or, virtually, in the speakers of an MP3 player. This observation accords broadly with views on (non-vocal/texted) music from the early-romantic period,

which celebrated it precisely because it lacked the conceptual precision of language and instead communicated more generalized, holistic phenomena. For E.T.A. Hoffmann (1776–1822), author of perhaps the most celebrated of such statements (Chantler, 2006), instrumental music

is the most romantic of all the arts – one might almost say, the only genuinely romantic one – for its sole subject is the infinite. The lyre of Orpheus opened the portals of Orcus – music discloses to man an unknown realm, a world that has nothing in common with the external sensual world that surrounds him, a world in which he leaves behind him all definite feelings [and concepts] to surrender himself to an inexpressible longing [*Sehnsucht*] (in Strunk, Treitler, & Solie, 1998, p. 151).

This is not to argue that music is a “universal language” (even though there are clearly certain “musical universals” resulting from gestalt and other perceptual-cognitive constraints (Lerdahl, 1992; Velardo, 2014)); but it appears likely that whereas we can glean very little linguistic information from speakers of languages with which we are unfamiliar, the music of other cultures often speaks to us directly and powerfully, despite its initial strangeness to us and our unfamiliarity with the details of its syntactic and expressive conventions. Moreover, while we might be oblivious to the grammatical structure of an unfamiliar language, we can discern a good deal of emotional information from its specifically musical elements – from the Hm-mmm-derived intonation of the speaker in conjunction with their facial expressions and body language. In such situations, we are transported back to the world of our hominin ancestors and compelled to activate our capacity to engage with the holistic, the manipulative, the multi-modal, the musical and – perhaps most important – the memetic.

## 7. References

- Adkins, M. (2009). The Application of Memetic Analysis to Electroacoustic Music. *Sonic Ideas*, 1(2), 34–41. Retrieved from <http://www.ems-network.org/ems08/papers/adkins.pdf>
- Agawu, V. K. (1991). *Playing with Signs: A Semiotic Interpretation of Classic Music*. Princeton, NJ: Princeton University Press.
- Aiello, L. C., & Dunbar, R. I. M. (1993). Neocortex Size, Group Size, and the Evolution of Language. *Current Anthropology*, 34, 184–193.
- Allanbrook, W. J. (1992). Two Threads through the Labyrinth: Topic and Process in the First Movements of K. 332 and K. 333. In W. J. Allanbrook, J. M. Levy, & W. P. Mahrt (Eds.), *Convention in Eighteenth- and Nineteenth-Century Music: Essays in Honor of Leonard G. Ratner* (pp. 125–171). Stuyvesant, NY: Pendragon Press.
- Bickerton, D. (2003). Symbol and Structure: A Comprehensive Framework for Language Evolution. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 77–93). Oxford: Oxford University Press.
- Blackmore, S. J. (1999). *The Meme Machine*. Oxford: Oxford University Press.
- Blackmore, S. J. (2005). *Consciousness: A Very Short Introduction*. Oxford: Oxford University Press.
- Boas, H. C., & Sag, I. A. (Eds.). (2012). *Sign-Based Construction Grammar*. Stanford, CA: Center for the Study of Language and Information.
- Bohlman, P. V. (2002). *World Music: A Very Short Introduction*. Oxford: Oxford University Press.
- Bonds, M. E. (1991). *Wordless Rhetoric: Musical Form and the Metaphor of the Oration*. Cambridge, MA: Harvard University Press.

- Braitenberg, V., & Braitenberg, C. (1979). Geometry of Orientation Columns in the Visual Cortex. *Biological Cybernetics*, 33, 179–186.
- Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, 23(10), 2791–2803. doi:10.1111/j.1460-9568.2006.04785.x
- Burak, Y., & Fiete, I. R. (2009). Accurate Path Integration in Continuous Attractor Network Models of Grid Cells. *PLoS Computational Biology*, 5(2), e1000291. doi:10.1371/journal.pcbi.1000291
- Byros, V. (2009). Towards an “Archaeology” of Hearing: Schemata and Eighteenth-Century Consciousness. *Musica Humana*, 1–2, 235–306.
- Calvin, W. H. (1996). *The Cerebral Code: Thinking a Thought in the Mosaics of the Mind*. Cambridge, MA: MIT Press.
- Cambouropoulos, E. (2001). Melodic Cue Abstraction, Similarity, and Category Formation: A Formal Model. *Music Perception*, 18(3), 347–370.
- Caplin, W. E. (2005). On the Relation of Musical Topoi to Formal Function. *Eighteenth-Century Music*, 2, 113–124.
- Carroll, S. B. (2003). Genetics and the Making of Homo Sapiens. *Nature*, 422(6934), 849–857.
- Carruthers, P. (2002). The Cognitive Functions of Language. *Behavioral and Brain Sciences*, 25, 657–726. doi:10.1017/S0140525X02000122
- Chantler, A. (2006). *E.T.A. Hoffmann’s Musical Aesthetics*. Aldershot: Ashgate.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Clegg, M. (2012). The Evolution of the Human Vocal Tract: Specialized for Speech. In *Music, Language, and Human Evolution* (pp. 58–80). Oxford: Oxford University Press.
- Cooke, D. (1968). *The Language of Music*. Oxford: Oxford University Press.
- Cross, I., & Tolbert, E. (2009). Music and Meaning. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford Handbook of Music Psychology* (pp. 24–34). Oxford: Oxford University Press.
- Crystal, D. (Ed.). (2003). *The Cambridge Encyclopedia of The English Language* (2nd ed.). Cambridge: Cambridge University Press.
- Darwin, C. (2008). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. (G. Beer, Ed.). Oxford: Oxford University Press.
- Dawkins, R. (1983). Universal Darwinism. In D. S. Bendall (Ed.), *Evolution from Molecules to Men* (pp. 403–425). Cambridge: Cambridge University Press.
- Dawkins, R. (1989). *The Selfish Gene* (2nd ed.). Oxford: Oxford University Press.
- Dawkins, R. (1991). *The Blind Watchmaker*. London: Penguin.

- Deliège, I. (2000). Listening to a Piece of Music: A Schematization Process based on Abstracted Surface Cues. In D. Greer (Ed.), *Musicology and Sister Disciplines: Past, Present, Future. Proceedings of the 16th International Congress of the International Musicological Society, London, 1997* (pp. 71–87). Oxford: Oxford University Press.
- Dennett, D. C. (1993). *Consciousness Explained*. London: Penguin.
- Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. London: Penguin.
- Deutsch, D. (1999). Grouping Mechanisms in Music. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 299–348). San Diego, CA: Academic Press.
- Dissanayake, E. (2000). Antecedents of the Temporal Arts in Early Mother-Infant Interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 389–410). Cambridge, MA: MIT Press.
- Distin, K. (2005). *The Selfish Meme: A Critical Reassessment*. Cambridge: Cambridge University Press.
- Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for Grid Cells in a Human Memory Network. *Nature*, *463*, 657–661. Retrieved from <http://dx.doi.org/10.1038/nature08704>
- Durham, W. H. (1991). *Coevolution: Genes, Culture, and Human Diversity*. Stanford: Stanford University Press.
- Durrell, M., Kohl, K., & Loftus, G. (2002). *Essential German Grammar*. London: Hodder Arnold.
- Fernando, C. T., Szathmáry, E., & Husbands, P. (2012). Selectionist and Evolutionary Approaches to Brain Function: A Critical Appraisal. *Frontiers in Computational Neuroscience*, *6*(24). doi:10.3389/fncom.2012.00024
- Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Foley, R. A. (2012). Music and Mosaics: The Evolution of Human Abilities. In *Music, Language, and Human Evolution* (pp. 31–57). Oxford: Oxford University Press.
- Fuhs, M. C., & Touretzky, D. S. (2006). A Spin Glass Model of Path Integration in Rat Medial Entorhinal Cortex. *The Journal of Neuroscience*, *26*(16), 4266–4276.
- Gamble, C. (2012). When the Words Dry Up: Music and Material Metaphors Half a Million Years Ago. In N. Bannan (Ed.), *Music, Language, and Human Evolution* (pp. 81–106). Oxford: Oxford University Press.
- Gjerdingen, R. O. (2007). *Music in the Galant Style*. New York: Oxford University Press.
- Goldberg, A. E. (2003). Constructions: A New Theoretical Approach to Language. *Trends in Cognitive Sciences*, *7*(5), 219–224. doi:10.1016/S1364-6613(03)00080-9
- Goldberg, A. E. (2013). Constructionist Approaches to Language. In T. Hoffmann & G. Trousdale (Eds.), *Handbook of Construction Grammar* (pp. 15–31). Oxford: Oxford University Press.

- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.
- Jan, S. B. (n.d.). Memetic Perspectives on the Evolution of Tonal Systems, under review.
- Jan, S. B. (2007). *The Memetics of Music: A Neo-Darwinian View of Musical Structure and Culture*. Aldershot: Ashgate.
- Jan, S. B. (2010). Memesatz contra Ursatz: Memetic Perspectives on the Aetiology and Evolution of Musical Structure. *Musicae Scientiae*, 14(1), 3–50.
- Jan, S. B. (2011). Music, Memory, and Memes in the Light of Calvinian Neuroscience. *Music Theory Online*, 17/2. Retrieved from <http://www.mtosmt.org/issues/mto.11.17.2/mto.11.17.2.jan.html>
- Jan, S. B. (2013). Using Galant Schemata as Evidence for Universal Darwinism. *Interdisciplinary Science Reviews*, 38(2), 149–168.
- Kirby, S. (2001). Spontaneous Evolution of Linguistic Structure: An Iterated Learning Model of the Emergence of Regularity and Irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2), 102–110.
- Kirby, S. (2007). The Evolution of Language. In R. I. M. Dunbar & L. Barrett (Eds.), *Oxford Handbook of Evolutionary Psychology* (pp. 669–681). Oxford: Oxford University Press.
- Kirby, S. (2013). Transitions: The Evolution of Linguistic Replicators. In P. M. Binder & K. Smith (Eds.), *The Language Phenomenon: Human Communication from Milliseconds to Millennia* (pp. 121–138). Berlin and Heidelberg: Springer.
- Kramer, L. (2002). *Musical Meaning: Toward a Critical History*. Berkeley and Los Angeles: University of California Press.
- Leman, M. (1995). *Music and Schema Theory: Cognitive Foundations of Systematic Musicology*. Berlin and Heidelberg: Springer.
- Leman, M. (2008). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA and London: MIT Press.
- Leng, X., & Shaw, G. L. (1991). Toward a Neural Theory of Higher Brain Function Using Music as a Window. *Concepts in Neuroscience*, 2, 229–258.
- Leng, X., Wright, E. L., & Shaw, G. L. (1990). Coding of Musical Structure and the Trion Model of Cortex. *Music Perception*, 8, 49–62.
- Lerdahl, F. (1992). Cognitive Constraints on Compositional Systems. *Contemporary Music Review*, 6(2), 97–121. doi:10.1080/07494469200640161
- Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Lynch, A. (1998). Units, Events and Dynamics in Memetic Evolution. *Journal of Memetics – Evolutionary Models of Information Transmission*, 2. Retrieved from [http://jomemit.cfpm.org/1998/vol2/lynch\\_a.html](http://jomemit.cfpm.org/1998/vol2/lynch_a.html)

- Matyja, J. R., & Schiavio, A. (2013). Enactive Music Cognition: Background and Research Themes. *Constructivist Foundations*, 8(3), 351–357. Retrieved from <http://www.univie.ac.at/constructivism/journal/8/3/351.matyja>
- Merker, B. (2002). Music: The Missing Humboldt System. *Musicae Scientiae*, 6, 3–21.
- Merker, B. (2012). The Vocal Learning Constellation: Imitation, Ritual Culture, Encephalization. In N. Bannan (Ed.), *Music, Language, and Human Evolution* (pp. 215–260). Oxford: Oxford University Press.
- Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.
- Meyer, L. B. (1996). *Style and Music: Theory, History, and Ideology*. Chicago: University of Chicago Press.
- Mhatre, H., Gorchetchnikov, A., & Grossberg, S. (2012). Grid Cell Hexagonal Patterns formed by Fast Self-Organized Learning within Entorhinal Cortex. *Hippocampus*, 22(2), 320–334. doi:10.1002/hipo.20901
- Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63(2), 81–97.
- Miller, J., & Van Loon, B. (2010). *Introducing Darwin: A Graphic Guide*. London: Icon Books.
- Mithen, S. (2006). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld & Nicolson.
- Monelle, R. (2006). *The Musical Topic: Hunt, Military and Pastoral*. Bloomington, IN: Indiana University Press.
- Morley, I. (2012). Hominin Physiological Evolution and the Emergence of Musical Capacities. In N. Bannan (Ed.), *Music, Language, and Human Evolution* (pp. 109–141). Oxford: Oxford University Press.
- Mountcastle, V. B. (1978). An Organizing Principle for Cerebral Function: The Unit Module and the Distributed System. In G. M. Edelman & V. B. Mountcastle (Eds.), *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function* (pp. 7–50). Cambridge, MA: MIT Press.
- Narmour, E. (1989). The “Genetic Code” of Melody: Cognitive Structures Generated by the Implication-Realization Model. *Contemporary Music Review*, 4(1), 45–63.
- Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press.
- Nattiez, J.-J. (1990). *Music and Discourse: Toward a Semiology of Music*. (C. Abbate, Trans.). Princeton, NJ: Princeton University Press.
- Patel, A. D. (2008). *Music, Language, and the Brain*. New York: Oxford University Press.
- Pinker, S. (1997). *How the Mind Works*. New York: Norton.

- Ratner, L. G. (1980). *Classic Music: Expression, Form, and Style*. New York: Schirmer.
- Ratner, L. G. (1991). Topical Content in Mozart's Keyboard Sonatas. *Early Music*, 19, 615–619.
- Reichl, L., Heide, D., Löwel, S., Crowley, J. C., Kaschube, M., & Wolf, F. (2012a). Coordinated Optimization of Visual Cortical Maps (I) Symmetry-Based Analysis. *PLoS Computational Biology*, 8(11), e1002466. doi:10.1371/journal.pcbi.1002466
- Reichl, L., Heide, D., Löwel, S., Crowley, J. C., Kaschube, M., & Wolf, F. (2012b). Coordinated Optimization of Visual Cortical Maps (II) Numerical Studies. *PLoS Computational Biology*, 8(11), e1002756. doi:10.1371/journal.pcbi.1002756
- Rice, T. J. (1997). *Joyce, Chaos, and Complexity*. Urbana and Chicago: University of Illinois Press.
- Schenker, H. (1979). *Free Composition*. (E. Oster, Ed.). New York: Longman.
- Scott-Phillips, T. C., & Kirby, S. (2010). Language Evolution in the Laboratory. *Trends in Cognitive Sciences*, 14, 411–417.
- Scruton, R. (1997). *The Aesthetics of Music*. New York: Oxford University Press.
- Shapiro, L. A. (2011). *Embodied Cognition*. London: Routledge.
- Snyder, B. (2000). *Music and Memory: An Introduction*. Cambridge, MA: MIT Press.
- Snyder, B. (2009). Memory for Music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford Handbook of Music Psychology* (pp. 107–117). Oxford: Oxford University Press.
- Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B., & Moser, E. I. (2012). The Entorhinal Grid Map is Discretized. *Nature*, 492, 72–78. Retrieved from <http://dx.doi.org/10.1038/nature11649>
- Strunk, W. O., Treitler, L., & Solie, R. A. (Eds.). (1998). *Source Readings in Music History: The Nineteenth Century* (Vol. 6). New York: Norton.
- Tagg, P. (1999). Introductory Notes to the Semiotics of Music, Version 3. Retrieved from <http://www.tagg.org/xpdfs/semiotug.pdf>
- Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press.
- Tolbert, E. (2001). Music and Meaning: An Evolutionary Story. *Psychology of Music*, 29, 84–94. doi:10.1177/0305735601291006
- Truss, L. (2003). *Eats, Shoots and Leaves: The Zero Tolerance Approach to Punctuation*. London: Profile.
- Velardo, V. (2014). The Sound/Music Dilemma: Why all music is sound but only some sounds are music. In *The Sound Ambiguity Conference*.
- Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings*. Cambridge, MA: MIT Press.

Wray, A. (1998). Protolanguage as a Holistic System for Social Interaction. *Language and Communication*, 18(1), 47–67.

Zatorre, R. J. (2003). Neural Specializations for Tonal Processing. In I. Peretz & R. J. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 231–246). Oxford: Oxford University Press.

## 8. Notes

---

<sup>1</sup> I am grateful to Lesley Jeffries, Valerio Velardo and the anonymous reviewer for their comments on earlier versions of this article.

<sup>2</sup> This separation was encouraged by the prohibition by the Société de Linguistique de Paris at its inception in 1866 of any discussion of the origins of language (Mithen, 2006, p. 1).

<sup>3</sup> As with musemes and lexemes, it seems reasonable to suggest that protemes, as self-contained units of information, were subject to evolutionary pressures and so, for a good deal of early hominin evolution, all three types formed a fuzzily overlapping and coexisting group of replicators.

<sup>4</sup> He nevertheless acknowledges the importance of the co-evolutionary relationship between biological and cultural forces in language evolution (Kirby, 2013, p. 136).

<sup>5</sup> To tone languages, IDS, and mantras one might tentatively add some forms of electroacoustic music. Several works in this medium lack clear segmental articulation, although they are not entirely beyond a memetic reading when inter-opus cross-stream mapping is evident (Adkins, 2009). Such music is clearly an analogue, not a homologue, of Hmmmmm, but might be predicted broadly to follow the course taken by Hmmmmm in its future evolutionary history and therefore afford evidence in support of the Wray/Mithen/Kirby hypothesis outlined here.

<sup>6</sup> I adopt Carruthers' convention here of using small capitals for concepts in mentalese and italics for internalized and vocalized language utterances.

<sup>7</sup> Structures located at the background, middleground and foreground layers are somatic; those elsewhere are extrasomatic. This hierarchic representation (after (Schenker, 1979)) is for expository clarity and is not intended to represent the topography of these functions in the brain, insofar as this is known.

<sup>8</sup> For the sake of expository clarity, the discussion suggest an element of unidirectionality; but in reality (and as implied by the double-headed arrows) it seems more likely that continuous bi-directional feedback loops connect structures at all three levels.

<sup>9</sup> Blackmore argues that consciousness presupposes a theory of mind and the associated capacity to ask "am I conscious now?" (Blackmore, 2005, p. 27).

<sup>10</sup> Morley notes the close coordination between language and motor centres in the brain (Morley, 2012, pp. 128–130), suggesting a strongly embodied/enactive aspect to musical and linguistic perception and production (Leman, 2008; Shapiro, 2011; Matyja & Schiavio, 2013).

<sup>11</sup> While such differentiation appears to have characterized human phylogeny, it is important to note that our ontogeny – the development of linguistic and musical competencies in individuals – might not necessarily rely upon domain-specific processes (Patel, 2008, p. 77).

<sup>12</sup> Many would argue that music has a semantic as well as an affective dimension (see Nattiez, 1990; Scruton, 1997; Kramer, 2002). What I am arguing here is that the mechanism by which this operates is parallel with that operating in language.

<sup>13</sup> After (Jan, 2007, p. 104, Table 3.1); the associated discussion is an extension of this earlier material.

<sup>14</sup> For clarity, Figure 2 ignores the motor-control memes which govern the muscular actions engendering writing, speaking and the production of musical sounds, many of which are learned as "implicit memory" and might be regarded as memes (Snyder, 2000, pp. 72–74).

<sup>15</sup> The term "interpretants" is Charles Sanders Peirce's (Nattiez, 1990, pp. 5–6). In Gottlob Frege's terminology, it aligns with the "sense" which qualifies and mediates the relationship between a term (a signifier/museme/lexeme) and its reference (a signified/object or concept) (Cross & Tolbert, 2009, p. 25).

<sup>16</sup> In language *I*, *GI* and *II* give rise to an essentially unary product: the concept is effectively inseparable from either its *I* or its *GI* or *II* manifestation, as symbolized by the curved brackets in column 4 of Figure 2 i a/b. In music, however, a separation is maintained, because *Gm* and *m* give rise to separate products – the notation (*Gm*) and, separately, the sounds which the notation motivates and regulates (*m*). Thus, unlike language, these two musical replicators preserve the level-two signifier-signified dualism at the phemotypic level.

<sup>17</sup> This is a general phenomenon in evolution, primarily observable in the inability of two species with a common ancestor to interbreed after a certain period of separate development has elapsed.