# Music, language and the signalling of cognitive ability: An empirical investigation

Keelin M. Murray

A Thesis submitted for the degree of Doctor of Philosophy

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

2013

# DECLARATION

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

*Keelin M. Murray*

# ABSTRACT

First systematically discussed by Darwin (1871), theories of a musical precursor to language have seen a revival in recent years, with researchers such as Tecumseh Fitch, Stephen Brown, and Stephen Mithen invigorating the field. The view that language and music have an evolutionary relationship has been discussed in light of recent comparative, musicological, and biological findings. However, little empirical data have been presented to support such theories. This thesis aims to address this oversight, by presenting a novel experimental paradigm, which tests the prediction of a novel hypothesis for the evolution of language from a musical precursor. The aim of this thesis is to encourage discussion and provide a framework for the empirical investigation of music's role in the evolution of language.

As a first step to addressing this relative dearth of empirical research, a hypothesis is outlined which describes a stable system of signalling cognitive ability through the transmission of culturally-learned, complex, music-like sequences. This is not hypothesised to have been semantically meaningful, rather a system which supported the honest transmission of information about the abilities of potential allies. Such a learned sequential precursor (LSP) to language would require both increased cognitive capacity and an investment of time and energy in learning. These requirements ensured the honesty of signalling, and so perceivers of the LSP could use it as a reliable indicator of the cognitive ability of producers. This was a necessary stage in evolution, prior to protolanguage, in which individuals exhibited a complex learned, culturally-transmitted, music-like signalling system. Such a learned sequential precursor may have arisen through a pressure for the reliable indication of cognitive ability, brought about by environmental and social changes with the advent of Homo erectus. These social changes included a new urge to cooperate, and so this precursor is proposed to have emerged and developed through collaborative partner choice. Perceivers of the system used cues within the musical sequences in order to determine the quality of a producer as a collaborative partner.

Empirical tests are presented, which support the hypothesised LSP. The first study tested the complexity aspect of the hypothesis, asking participants to rate complex and non-complex pieces of music according to how much they liked the piece, how familiar it sounded, how attractive and intelligent they found the person who created it, and how likely they were to choose to collaborate with this individual. It was found that complexity was preferred under all measures but one, that of familiarity. The second, main, study predicted that a correlation should be found between measures of cognitive ability that are relevant to musical learning (processing speed and intelligence) and measures of musical learning (ability to replicate and recall target pieces, and make creative pieces). This prediction was upheld, supporting the hypothesis that a learned sequential precursor could have acted as an honest signal of cognitive ability. No correlations were found between these abilities and a measure of physical quality, supporting the hypothesis that this system may have undergone social selection. The third study further tested the question of selection and choice, predicting that collaborative partner choice was key to the selection of this learned sequential precursor. Raters were asked to rate the sexual or collaborative ability of performers of pieces of music, based solely on their musical output. This study has yielded interesting tendencies, but no statistical support of the hypothesis that collaborative partner choice was more important than mate choice in this system. Taken together, these empirical studies support the hypothesis of a musical, learned sequential system of signalling cognitive ability. At the moment, the question of the selection of this precursor remains open, with hopes that further studies can address this question. The methodology used here draws together approaches from birdsong research, evolutionary psychology, and musicological research, in an attempt to prompt further interdisciplinary investigation into the role of music in the evolution of language.

# ACKNOWLEDGEMENTS

This thesis would never have come together if it weren't for a vast number of people. In the interests of space, I will keep my acknowledgements short, but those who were involved in any small way should consider themselves thanked. First thanks must go to my supervisors, Simon Kirby and Lars Penke. They were the ideal pair, complementing each other's approaches and always keeping me on my toes. Mónica Tamariz was involved in my supervision in first year, and among many other things, encouraged me to think visually about tricky problems. Thanks to the LEC - I feel spoiled by having been part of such a supportive, varied group. This research benefitted greatly from discussion, argument, and challenge from all members of the group. My office mates in 1.15 have lightened the load with chat, gossip and cake. Thanks to Katie Overy for early musicological input, whilst Barbora Skarabela showed me that women in academia can and should support each other. Special thanks to Katie Keltie, Toni Noble, and Lynsey Buchanan, who dealt with all of my questions with skill and good grace. Eddie Dubourg and Ziggy Campbell gave invaluable help with musical stimuli and computer issues. Thanks to the ESRC for funding this research.

On a personal note, thanks must go to my closest friend Sarah Yabroff, for inspiring me in the last few months of writing up - I can't wait to see her fulfill her own PhD plans. Evonne Pasquill took on the brave task of living with a final year PhD student, and in doing so, kept me well fed and happy. Séan Roberts is my oldest Edinburgh friend, and has both inspired me daily and solved more problems for me than he ever should have. Manuela, Juan, Amanda and Márton helped me through the dark days. Finally, thanks to my family. Aoife and Shane gave me a bolthole to escape to, and Aoife has been the best sister, counsellor, and friend that I could have hoped for. My parents Margaret and Frank made all of this possible, with their constant support in every way. They have a faith in my abilities that at times I haven't had myself. It is thanks to them that I can present this thesis to the reader.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

Language permeates every aspect of our lives, and we only become aware of this when language fails us - when in a foreign country, for example. Furthermore, language is acquired easily and naturally by children, with no need for explicit tuition (Ramscar and Yarlett, 2007). Music seems also to be a uniquely human trait (McDermott and Hauser, 2005) - although animals, in particular birds, do exhibit song or call (Seyfarth et al., 1980; Catchpole and Slater, 1995; Slater, 2000; Spencer et al., 2003), this is a communication system, quite different in form and function from both language and music. Music, like language, is seen in all known human cultures (Nettl, 2000), and permeates these cultures entirely. Musicality is seen early in childhood development (Malloch, 1999; Welch, 1998; Trainor and Unrau, 2012; Nakata and Trehub, 2004), and from an early age, children can engage in musical interaction with others (Trevarthen, 1999), with two-month old infants able to recognise distinct melodies (Plantinga and Trainor, 2009). Music and language both consist of hierarchically arranged elements which can be recombined to make a novel utterance or melody (Lerdahl and Jackendoff, 1983). Indeed, they appear to share some of the same processing systems (Koelsch et al., 2002, 2004; Federenko et al., 2009), and behave similarly in syntactic processing (Maess et al., 2001). For these reasons, and others which will be addressed in this thesis, it seems appropriate to study music and language side-by-side - in particular, the evolution of both faculties. In this thesis, I aim to address the question of whether music played a role in the evolution of language. I follow the tradition of many authors, most notably Charles Darwin, who first laid out a scientific approach to music's role in language evolution. However, I will take this investigation one step further by adding to ongoing experimental work, testing theories such as Darwin's - can we empirically test whether language may have evolved out of music?

The evolution of language is a field which has seen great interest in recent decades. Historically (wrongly) relegated to "armchair linguists" and speculation, we have seen an explosion in mathematical modelling (Nowak and Krakauer, 1999), computer simulations (Kirby, 2001; Ritchie and Kirby, 2006), and experimental investigation (Kirby et al., 2008; Roberts, 2010; Verhoef et al., 2012; Simner et al., 2010) since the late 1990s. Psychology, on the other hand, has a decidedly long and rich empirical basis, and it is from this that empiricists in language evolution should and have taken inspiration. Further, one cannot overlook musicological approaches, which have also taken an experimental approach of late (Kippen and Bel, 1992; Kirschner and Tomasello, 2010). This thesis will draw on all of these aspects, to attempt to lay a foundation for the testing of musical protolanguage theories.

In this thesis I will propose and empirically test a novel theory of the evolution of language from music. This theory is focused upon what will be called the *Learned Sequential Precursor* (LSP). This is proposed as a stable system of signalling cognitive ability through the transmission of culturally-learned, complex, music-like sequences. This system required and drew upon both increased cognitive capacity and an investment in learning. These requirements ensured the honesty of signals, and so perceivers of this learned sequential precursor could use it as a reliable indicator of the cognitive ability of producers. This was a necessary stage in evolution, prior to protolanguage, in which a learned, culturally-transmitted music-like system was in use. This precursor arose due to a pressure for the honest signalling of cognitive ability, itself a result of changing environmental and social pressures associated with Homo erectus. A figure outlining the major transitions in this system can be seen in Figure 1.1 on page 3.

The second chapter will discuss the LSP in detail, and support from evolutionary psychology, birdsong research, and musicology will be presented. I will address human evolution in chapter three, in order to situate in time this musical signalling system. Furthermore, the mechanisms through which this signalling system evolved to language will be discussed, and hypotheses about what this means for language outlined.

Chapter four will outline two studies which aim to address the role of music in a precursor to language. The first will look at complexity, and aims to cast light on the proposed role of complexity in a learned signalling system. The second, main, study of this thesis aims to test the predictions of the learned sequential precursor hypothesis. Chapter five will present another study, into the proposed mechanisms leading to the selection of the hypothesised musical precur-

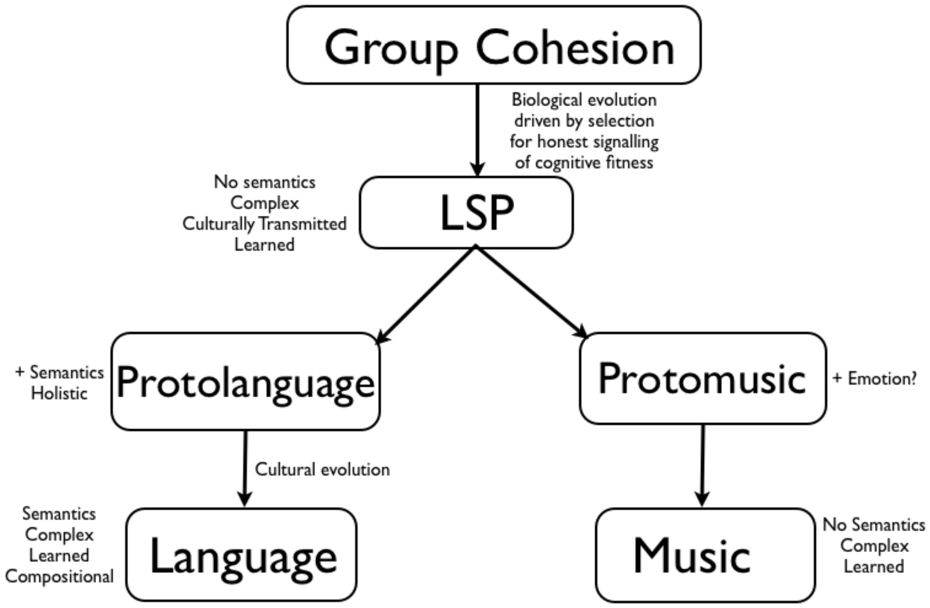


Figure 1.1: Major transitions in language evolution
Major transitions in language evolution, with the LSP an intermediate stage between a posited rhythmic group cohesion system, and protolanguage

sor. This experimental work will hopefully provide a platform for future research into musical protolanguage. It is hoped that this thesis will provide convincing evidence that theories of the origins of language from music *can* be empirically tested, if we are prepared to be creative and work in an interdisciplinary fashion. First, however, this present chapter will outline existing theories of musical protolanguage, as well as separate theories which address the evolution of music.

## 1.1 Overview of Musical Protolanguage theories

Theodore Adorno was not the first to note the very tangible similarities between music and language. However, his suggestion that "in our day the relationship between music and language has become critical" (Adorno, 1956, 1) spelled out in explicit terms the need for researchers in many fields to consider music and language side by side. Indeed, recently this idea was expounded upon by Masataka (2009, 11): "the fact that both are human universals, have phrase structure, and entail learning and cultural transmission, suggests that any theory of the evolution of language will have implications for the evolution of music, and vice versa". This thesis will mirror that approach, in attempting to address the evolution of both music and language, whilst focusing in the main on language. Let us first address some views regarding the evolution of these two, uniquely human, traits. Musical protolanguage theories are relatively few, with the majority of theoretical analyses having been conducted in the last two decades.

One of the first scholars to discuss some form of musical precursor to language was Lord Monboddo (James Burnett), who, in 1774, wrote that "there is an ingenious man, an acquaintance of mine... who conjectures, that the first language amongst men was *music*, and that before our ideas were expressed by articulate sound, they were communicated by tones... musical tones, which we first learned by imitation of the birds" (Burnett, 1774, 469-470). He states that "it is natural enough to think of applying the variation of tones to a purpose of utility as well as pleasure" (Burnett, 1774, 470). Whilst this is an interesting mention of a musical precursor to language, which is attributed by Monboddo to "Dr Blacklock of Edinburgh; a person of great genius, and wonderful learning" (Burnett, 1774, 469), Monboddo himself concludes that he is unconvinced by a "language of singing merely" (Burnett, 1774, 473), and so this concludes his interest in this area.

The first fully-fledged, in-depth theory which proposed a role for music in the evolution of language came from Charles Darwin, in his hugely influential book *The Descent of Man and Selection in Relation to Sex* (Darwin, 1871a). His theory has had far-reaching effects - in the late 19th and early 20th centuries, Otto Jespersen (1895, 1922) attempted to expand upon the theory outlined by Darwin, proposing the evolution of language from this musical (sung) system. In recent years, the musical protolanguage theory (so-called by modern theorists) has been revisited, sometimes lamentably without adequate reference to Darwin's original hypothesis. The latest discussion of musical protolanguage comes from Tecumseh Fitch (2010), who lauds Darwin's hypothesis, while proposing the term "prosodic protolanguage". This section will provide an overview of the main theories in this field, and will highlight areas of interest or inspiration for the main hypothesis, to be outlined in chapter two. The theories will be presented in chronological order, in order to highlight the evolution of ideas from Darwin's theory through time.

### 1.1.1 DARWIN

Charles Darwin, in 1871's *The Descent of Man*, addressed the origin and evolution of language in a short, eleven-page section. He wrote

> "That which distinguishes man from the lower animals is not the understanding of articulate sounds ... it is not the mere articulation which is our distinguishing character ... nor is it the mere capacity of connecting definite sounds with definite ideas ... [it is the] almost infinitely larger power of associating together the most diversified sounds and ideas, and this obviously depends on the high development of his mental powers" (Darwin, 1871a, 54).

Stating that "the sounds uttered by birds offer in several respects the nearest analogy to language", Darwin suggested a largely musical communication system in primeval man - our ancestor "probably first used his voice in producing true musical cadences, that is in singing" (Darwin, 1871a, 56). This system was proposed to have been "especially exerted during the courtship of the sexes - would have expressed various emotions, such as love, jealousy, triumph, and would have served as a challenge to rivals" (Darwin, 1871a, 56). This was a very ancient system, as Darwin (1871b, 334) writes; "whether or not the half-human progenitors of man possessed, like the before-mentioned gibbon, the capacity of producing, and no doubt of appreciating, musical notes, we have every reason to believe that man possessed these faculties at a very remote period, for singing and music are extremely ancient arts".

Darwin discussed how this may have come to be: "some unusually wise ape-like animals should have thought of imitating the growl of a beast of prey... it is probable that the imitation of musical cries by articulate sounds may have given rise to words expressive of various complex emotions" (Darwin, 1871a, 57). Thus, hominins adapted the calls in their environment, creating a music-like, cadential system, to increase their chance of mating, and to express emotions and dispute territorial issues. With continued use of this communication system, Darwin posits that the vocal organs would have been perfected. However, he stipulates that this merely physiological change was overshadowed by the mental development necessary for "even the most imperfect form of speech" to come into existence: "we may confidently believe that the continued use and advancement of this power [imperfect form of speech] would have reacted on the mind itself, by enabling and encouraging it to carry on long trains of thought" (Darwin, 1871a, 57). Darwin hypothesized a sexual selection pressure for language and music: "we may assume that musical tones and rhythm were used by our half-human ancestors during the season of courtship. . . from the deeply-laid principle of inherited associations, musical tones in this case would be likely to call up vaguely and indefinitely to strong emotions of a long-past age" (Darwin, 1871b, 336-7). He went further, to say "women are generally thought to possess sweeter voices than men, and as far as this serves as any guide we may infer that they first acquired musical powers in order to attract the other sex" (Darwin, 1871b, 337).

Darwin strongly argues against the concept that music evolved from speech, in a section added to the second edition of his treatise, stating;

> "As we have every reason to suppose that articulate speech is one of the latest, as it certainly is the highest, of the arts acquired by man, and as the instinctive power of producing musical notes and rhythms is developed low down in the animal series, it would be altogether opposed to the principle of evolution, if we were to admit that man's musical capacity has been developed from the tones used in impassioned speech. We must suppose that the rhythms and cadences of oratory are derived from previously developed musical powers . . . musical sounds afforded one of the bases for the development of language" (Darwin, 1874, 572).

In essence, Darwin posed a hypothesis in which our early ancestors used a musical system to communicate and mediate sexual and territorial issues. This system would have undergone sexual selection; "the vocal organs were primarily used and perfected in relation to the propagation of the species" (Darwin, 1871b, 330), with mental development and vocal improvement eventually leading to increased cognition and a speech-like linguistic system. Music remains as a fossil of this earlier protomusical system (Fitch, 2012, 89).

On its publication, Darwin's *The Descent of Man* was publicly reviled and dismissed. Darwin was attacked on two fronts; scientifically, Charles Alexander Hodge wrote that the theory of evolution was too speculative, and that therefore all conclusions could be ignored, as Darwin had failed to develop a theory based on "truth" (Moore, 1981; Russett, 1976, 203-4). Darwin was also attacked on a religious front, by many authors including Enoch Fitch Burr, who wrote that evolution was "founded by atheism, claimed by atheism, supported by atheism, used exclusively in the interest of atheism" (Moore, 1981, 197-8). Furthermore, Darwinism was rejected as "a most gigantic hoax" (Moore, 1981; Gray, 1876, 329-330). A summary of the controversy created by Darwin's theory of evolution is available in Webb (1994). While we shall see that

Otto Jespersen (1895, 1922) did continue theorising about a musical predecessor to language in the years following Darwin's publication, it is only in recent years that Darwin's theory of protolanguage has seen a real revival. Although Darwin's discussion of language only spans eleven pages, the importance of his hypothesis cannot be overlooked. His hypothesis is remarkably modern and accessible and has (perhaps indirectly) influenced recent authors, some of whom will be discussed below.

### 1.1.2 JESPERSEN

Otto Jespersen was one of the first twentieth century writers to suggest a musical origin for language, drawing on contemporary theories of "primitive" societies. He begins his thesis of music's role in language evolution by stating that

> "It is a consequence of advancing civilisation that passion, or at least, the expression of passion, is moderated, and we must therefore conclude that the speech of uncivilised and primitive men was more passionately agitated than ours, more like music or song" (Jespersen, 1922, 424).

He writes that there must have been a time when all speech was song, or rather "when these two actions were not differentiated" (Jespersen, 1922, 424) . . . "our comparatively monotonous spoken language and our highly developed vocal music are differentiations of primitive utterances, which had more in them of the latter than of the former" (Jespersen, 1922, 436).

Jespersen's theory proposes that extended altriciality due to bipedalism, and social bonding conditioned early man for vocal play, stating for example "Language originated as play, and the organs of speech were first trained in the singing sport of idle hours" (Jespersen, 1922, 433), and language "began with half-musical unanalyzed expressions for individual beings and solitary events" (Jespersen, 1922, 441). He writes that the language of our ancestors was "like that ceaseless humming and cooing with which no thoughts are as yet connected" (Jespersen, 1922, 433). Jespersen follows Darwin in proposing sexual selection for this interaction system, stating that "thoughts were not the first things to press forward and crave for expression; emotions and instincts were more primitive and far more powerful...and among the emotions which were most powerful in eliciting outbursts of music and of song, love must be placed in the first rank" (Jespersen, 1922, 433).

Jespersen makes a valid point which must be kept in mind throughout this thesis - at no point, was this an *conscious* system - “Our remote ancestors had not the slightest idea that such a thing as communicating ideas and feeling to someone else was possible”. Furthermore, we must not confuse the “musical” aspect of many of these theories with modern music - as Jespersen writes “as is the art of writing to primitive painting, so is the art of speaking to primitive singing” (Jespersen, 1922, 437). Jespersen’s theory, though built upon archaic notions of “the primitive”, mirrors Darwin’s in choosing sexual selection as the key mechanism for the development of a musical precursor to language. Following Jespersen’s 1922 publication, a gap of almost 50 years developed before this theory was revisited, by musicologist Bruno Nettl.

### 1.1.3 NETTL

Bruno Nettl, renowned ethnomusicologist, developed a theory of the evolution of music and language out of “an undifferentiated method of communication . . . which was neither speech nor music but which possessed the three features that they hold in common; pitch, stress, and duration” (Nettl, 1969, 136). Nettl claims that this proposed method of communication didn’t have fixed pitch, nor did it have regular stress patterns or durational values, nor distinctions between vowels and consonants. The sounds produced were grunts, cries and wails, which sounded like neither music nor language, but which embodied pitch, stress and duration. From this system, both music and language evolved, through a “long, gradual stage of differentiation and specialization in culture” (Nettl, 1969, 137). During this differentiation, music took on fixed pitch and language took on vowels and consonants. So, Nettl’s theory takes three steps: the first consisted of undifferentiated communication, the second of differentiation between the two faculties, language and music, with music still in an elementary stage, and the third which consisted of the differentiation between various musical styles, his particular area of expertise. Nettle claims that the simplest primitive musical styles, such as litany-type forms (short phrases repeated with little or no variation) and ditonic scales (scales with only two notes), have a long period of development behind them. However, he asserts that what he calls “primitive” music is “a far cry from man’s earliest musical experiences, just as the simplest extant cultures cannot be assumed to resemble those of prehistoric times” (Nettl, 1969, 137) - here, Nettl echoes Jespersen (1922).

This is an interesting theory, obviously more music- than language-oriented. The three steps on the road to music could easily be adapted to three steps on the road to language (and different languages). Pitch stress and duration are deemed relevant here to a precursor to both music and language, and indeed both pitch and duration are key to musical learning, as defined in the later experimental section of this thesis (chapter four). Nettl's theory behaves as a stepping-stone, between the theories of Darwin and Jespersen in the late 19th and early 20th centuries, and modern theories, serving to introduce the concept of an "undifferentiated" system, which would be repeatedly revisited in the following decades.

### 1.1.4 LIVINGSTONE

Livingstone, in his 1973 paper "Did the Australopithecines sing?" lays out his theory of the evolution of language and music. Stating that singing is simpler than speech, Livingstone contends that man "could sing long before he could talk and that singing was in fact a prerequisite to speech and hence language" (Livingstone, 1973, 25). There "would have had to be a considerable amount of evolution from the chimpanzee call system to a learned set of personal and group songs" (Livingstone, 1973, 26). He claims, drawing on Haldane (1955), that "songs as group or personal names" may have served to open the call system. This adaptation to a "learned, open signal system of territorial songs preadapted the hominins to both speech and symbolising" (Livingstone, 1973, 26). He also states that early man probably made use of many gestures, whilst claiming that gestural theories alone can not explain the design features of language. Both these territorial songs and gestures evolved through some "specific impetus" (Livingstone, 1973, 26) to language. Livingstone espouses the uncontroversial view that the possession of language by some groups would have been a powerful isolating tool, and would have resulted in speciation. He claims that early Neanderthals fulfilled the requirements of his system - the rudimentary ritual and symbols - and that this adaptation to language may have allowed them to supercede groups of Homo erectus in Eurasia and Africa.

There are some issues to be addressed in Livingstone's theory. Initially, in laying out his evidence for his theory, Livingstone (1973, 25) states that "singing is a much simpler system than speech, with only pitch as a distinguishing factor". However, it is unclear exactly what he means by "distinguishing factor". If he means simply the ability to distinguish between songs, then surely dynamic and rhythmic structure both serve as such a factor? Furthermore, Livingstone states that “adaptation to this learned, open signal system of territorial songs preadapted the hominins to both speech and symbolising” (Livingstone, 1973, 26). Here, he seems to see symbolic behaviour and speech coming on stream at the same time. However, the ability to use symbols must appear before language can, precisely due to language’s symbolic nature. Does Livingstone mean that this adaptation preadapted hominins to *vocalisations* and symbolising? This may be a more prudent interpretation - alternatively Livingstone’s theory needs a more rigorous analysis.

### 1.1.5 LEVMAN

Another writer to suggest an evolutionary precursor to music and language is Bryan G. Levman. He contends that “language and music evolved out of a common ‘proto-faculty’, which was primarily musical in nature” (Levman, 1992, 147). Again, here we see the theory, first introduced by Nettl, of a *common* proto-faculty. Stating that both language and music evolved out of the “fundamental impulse of an organism to survive” (Levman, 1992, 150), for which hearing and vocalisation were indispensable, Levman writes that

> “The first language of humans was therefore a form of music/speech which they developed to help navigate in the environment and increase chances of survival. This primal language was much more akin to what we understand as music than it was to speech” (Levman, 1992, 150).

Levman states that double articulation (duality of patterning) evolved to express more complicated concepts than pitch variation alone could express, due to increased complexity of living. With this innovation, the ‘first language’ bifurcated into two evolutionary paths which, though related, remain distinct. Language grew “in the direction of greater and greater lexical specificity and syntactical sophistication, music concentrating on melodic and rhythmic refinement and combination” (Levman, 1992, 165). He goes on to say that pitch must have played “as important a role in humankind’s protolanguage as it does today in music” (Levman, 1992, 152), and that differences between music and speech such as rhythm, dynamics, and articulation, should not be viewed as a differences of kind, rather than of degree. Levman proposes that ‘this musical faculty [from which language evolved] is a direct development of the organism’s sonic perception/production capacities, and was selected by evolution as an important survival aid to assist the organism in its intra- and interspecies and environmental navigations" (Levman, 1992, 164).

### 1.1.6 BROWN

One of the most well-known modern theories of our early protolanguage is Steven Brown's proposed "musilanguage" stage, a common ancestor of both music and language. In his view, the many striking similarities between music and language are as a result of a "joint evolutionary precursor rather than from fortuitous parallelism or from one function begetting the other" (Brown, 2000, 271). This proposed musilanguage stage was a simple system with a repertoire of lexical-tonal units, which evolved to a "less simple system based on combinatorial arrangements of these lexical-tonal (and rhythmic) elements" (Brown, 2000, 290), with expressive phrasing such as meaning association, prosody, accent and stress in place. Both music and language could have evolved from this precursor, explaining the many shared features. Of course, any analogous or distinct features of either system can be explained by divergent evolution after this musilanguage stage. Brown suggests a Referential Emotive Vocalization (REV) system, similar to that seen in vervet monkeys, as the natural origin of the musilanguage system. Vervet monkeys have three distinct calls, which seem to correspond to snake, leopard, and eagle (Seyfarth et al., 1980). These calls elicit different, situation-appropriate responses in hearers. This REV has a dual acoustic nature, in which a given sound pattern (or call) has both emotive and referential meaning. In other words, each call may convey both referential - "there is a snake" and emotional - "you should be scared of this snake" meaning. After the musilanguage developed into both language and music, these two roles can be seen - language is sound as referential meaning, while music is sound as emotional meaning. Indeed, Brown sees language and music as opposite ends of a continuum, on which we can find such mixing of language and music as heightened speech, poetic discourse, word painting, and musical narration (Brown, 2000, 275). We might also suggest that such techniques as Sprechstimme, (an operatic technique which falls halfway between speech and song) and Recitative (a technique in which the singer adopts the rhythms of speech) could be situated on this continuum. This idea of a continuum between speech and song mirrors the work of List (1963). Indeed, Brown's discussion of a continuum might have been enriched by drawing on this literature.

In relation to the question of selection, Brown writes that

> “There is just too much about music making that reveals an essential role in group function to ignore the issue of multi-level selection. . . a principal function, if not the principal function, of music making is to promote group cooperation, coordination and cohesion” (Brown, 2000, 296).

Thus Brown’s hypothesis diverges from Darwin’s in this crucial factor. He sees the modern role of music as group cohesion as indicative of the prior role of his “musilanguage”. Brown proposes both group selection (Sober and Wilson, 1998) and cultural group selection (Boyd and Richerson, 1990), as they ‘offer great promise” in clarifying the cooperative, group based nature of music (Brown, 2000, 297). Group selection is an instance of natural selection operating between groups of organisms, rather than between individual organisms. It produces adaptations that benefit the group rather than the individual.

Brown attempts to address the striking similarities and shared features between music and language, by discussing five models, which he terms the *parallelism, binding, music outgrowth, language outgrowth*, and *musilanguage* models. As he defines them, the parallelism model proposes that music evolved from a protomusical precursor, and language evolved from a protolinguistic precursor, entirely independently, with no common form. The binding model proposes that language and music each evolved independently, but that binding mechanisms evolved which conferred linguistic properties onto music, and musical properties onto language. These two models do not consider any common ancestor. The next three, music outgrowth, language outgrowth, and musilanguage models, do posit a common predecessor. The music outgrowth model proposes that music evolved from a protolinguistic precursor, while the language outgrowth model proposes the opposite, that language evolved from a protomusical precursor. Brown provides evidence to allow him to put to one side the first four, before focusing on the fifth, musilanguage, model. By his analysis, the parallelism model is not parsimonious, whilst the binding model rests on “an overly dichotomous view of music and language” (Brown, 2000, 277), and is refuted by evidence from lesions in the brain which eliminate musical properties of speech while sparing those of music, and vice versa. Here, he refers to Borchgrevink (1991), providing evidence that selective anesthesia of the right hemisphere of the brain disrupts pitch in singing, but spares speech prosody. His argument for disregarding the two outgrowth models, music outgrowth and language outgrowth (which describes Darwin's theory), is that the musilanguage model "greatly simplifies thinking" (Brown, 2000, 277) about musical and linguistic origins. Brown claims that the musilanguage model avoids semantic ambiguities about what constitutes an ancestral musical or linguistic property, by taking as its starting point the common features of both, and referring to these features as musilinguistic. From this musilanguage stage, the distinct features of both music and language occurred later in human evolution.

However, Brown seems to be over-playing the importance of his terminology here. The distinctions he makes between, for example, the music outgrowth model and the musilanguage model are purely semantic. In order to "greatly simplify thinking" about the models, he has simply created a new model to add to the system - his lauded term "musilanguage" does nothing to avoid semantic ambiguities. As the definitions of "language" and "music' are not clearly defined here, and each take somewhat different roles in the models Brown discusses, it appears to be simply a semantic issue to distinguish between these five models.

### 1.1.7 Merker

It is telling that a collection entitled *The Origins of Music* (Wallin et al., 2000) should feature multiple discussions (Brown 2000, Merker 2000) of the role of music in the evolution of language. Merker (2000) adds to the debate with the hypothesis that synchronous chorusing was key to the evolution of humankind, and perhaps to the evolution of language. Noting that musical pulse is a "cardinal device" for coordination of individuals in a joint task or performance, Merker claims that "synchronous chorusing may have played a fundamental and hitherto unsuspected role in the process of hominin divergence from our common ancestor with the chimpanzee" (Merker, 2000, 315). Synchronous calling or chorusing does not appear in a vaccuum, and must be associated with a "motivational mechanism for mutual entrainment" (Merker, 2000, 318). Merker assumes that this mechanism was selected for in the course of divergence from our ancestors, and has been retained today "in the form of our propensity to join in and entrain to a repetitive beat" (Merker, 2000, 318). This synchronous chorusing is proposed to have been coordinated with the aid of "bodily movements derived from the repertoire of walking and running" (Merker, 2000, 319), in other words, dancing. All of this sums up to suggest that our hominin ancestors engaged in "a novel and unique social adaptation, namely a behavioural forum featuring synchronous singing and dancing on the part of a higher animal" (Merker, 2000, 320). Once this system was in place, it may have been affected by sexual selection in two fashions; one, migrating groups of females would choose between groups of chorusing males on the basis of their output, and two, individual males in the group would have been distinguished between in mate choice. We might, then, view this system as an example of individuals collaborating in order to maximise their own reproductive success. Merker's discussion of the role of synchronous chorusing in the evolution of language is deliberately speculative, but he does state that "behavioural capacities and biases based on perceptual, motivation, cognitive and motor mechanisms evolved for other purposes [may have been] so constituted as to supply essential foundations for human language (Merker, 2000, 323). Chorusing vocalisations may have come to be associated with specific elements of the world, as vocal signatures. This may have become a prototype for "generalised naming by distinctive, elaborate phrase patterns" (Merker, 2000, 322). Merker leaves us with a challenge, closing his proposal by stating "it is necessary to know whether or not we were in fact singing and dancing hominins before we became talking humans, and if so, whether and how long we might have been singing and dancing humans before we started to employ our cerebral equipment for referential language" (Merker, 2000, 323).

### 1.1.8 Miller

Geoffrey Miller is another advocate, following Darwin, of the sexual selection hypothesis for music and language. In an attempt to address the "apparently altruistic act of speaking", Miller (2001, 349) quickly dismisses the kinship and reciprocity hypotheses, stating that their predictions are not supported by real-life human interaction. Miller instead proposes a verbal courtship hypothesis, which, he states, is "the heart of human sexual selection" (Miller, 2001, 351). Using the analogy of the language of a teenage boy, Miller draws similarities between this boy and our ancestors - "poor vocal control, small vocabulary, uncertainty about conversational conventions, difficulty in finding phrases to express thoughts" (Miller, 2001, 352). However, Miller does not delve further into the origins of language, going no further than to state: "once the rudiments of language started to evolve, *for whatever reason*" (Miller, 2001, 352) - (emphasis added). This is an unfortunate approach to take, as further hypotheses for the origin of language can only encourage further investigation (theoretical and empirical) into the evolutionary mystery.

Furthermore, Miller's key argument for sexual selection, that "if natural selection has shaped human language for the efficient, cooperative communication of useful information, we would all speak this sort of 'early adolescent mumbled dialect" (Miller, 2001, 352) seems unwarranted. His support for this view is that "the boy's same-sex friends seem to demand little more than quiet, cryptic, grammatically degenerate mumbling" (Miller, 2001, 352). However, this seems somewhat of a just-so story not supported by rigorous scientific endeavour, if not insulting to many teenage boys.

So, Miller claims that sexual selection of language solves the problem of altruism, as speaking well would entail a sexual payoff - thus, language evolved, via sexual selection, to that which we have today. Miller takes an even bolder stance on the evolution of music, stating that "this hypothesis [of courtship] for human music is not only better supported by music's design features, but should be considered the evolutionary null hypothesis" (Miller, 2000, 330). These design features in his estimation include universality, complexity, and spontaneous development. Furthermore, Miller claims that certain features of music are products of sexual selection, going on to list its spontaneous practice and production "despite energetic costs and lack of survival utility" (Miller, 2000, 337). He also uses the examples of musicians avoiding sex before an important concert in order to keep a "sexual edge", to support his proposal that music is used conspicuously in courtship, and that its production declines after mating. Additionally, Miller states that

> "Public music production rockets upward after puberty, reaches its peak in young adulthood during the period of most intense courtship, and declines gradually with age and parenting demands. Musical tastes lead to strong assortative mating" (Miller, 2000, 337-8).

However, these pieces of evidence need not necessarily support a sexual selection hypothesis. Let us take its "spontaneous practice... despite energetic costs and lack of survival utility". This need not be as a result of sexual selection. This evidence would support other group-based hypotheses. Miller's example of musicians, like sportsmen, holding off from sex before an important performance could equally well be explained by the need for a good night's sleep before an important concert. Musicians don't want to be exhausted before the draining physical exercise involved in a musical performance. Miller's discussion of public music production coinciding with puberty and other life changes, could well be explained by just that - young adults, at puberty, are encountering new people, they are at a point of extreme change in their life - new school, new friends, increased freedom for example. This creates a state of flux, which largely stays in place until the early to mid twenties. Public music making need not be a sexually selected trait, but could be as a result of social group formation. The decline of music production in later years, noted by Miller, could be explained by the adult now being comfortable and settled in their social group. Thus, they have no need to impress other group members with their musical skill. Furthermore, I would suggest the high numbers of children who choose to drop out of music lessons in their teens as contradictory to Miller's hypothesis. Any music teacher could name many students whose musical performance hasn't survived the early teen years - as a former violin teacher, I can personally attest to this. How does Miller explain this? As for the claim that musical tastes lead to assortative mating, when looked at closely, this seems a leap. I would qualify this statement by stating that musical tastes lead to group formation, and thus, *by extension*, assortative mating. However, in these groups, music is unlikely to be the sole common factor. For example, fans of jazz music are likely to be of similar age, education status, and background (although of course exceptions do exist, we are generalizing here). The mating is not directly as a result of musical tastes, but rather, via group membership and shared interests.

Miller goes on to discuss why he believes that group selection models cannot viably explain music evolution (and, by extension, language evolution). He argues that the first error of the group approach is ideological - group selection can be seen as "kinder, gentler, more cooperative" (Miller, 2000, 351).

Miller is not alone in drawing on modern day musical behaviour to conclude that music must be a sexually selected trait. Slater (2000, 59) writes that "the fact that in many cultures singing (and in our own culture, composition) is predominantly a feature of young males. . . confirms that suspicion [of sexual selection]". He refers to Miller (2000) in support of this "fact". However, as seen above, Miller provides little tangible evidence outside of speculation about the sexual practices of famous guitar players. Until a large-scale analysis of real musicians and composers (and how is "musician" or "composer" defined - professional/ amateur?) is conducted, this point is null.

Recent work which aimed to support the sexual selection hypothesis for music (Charlton et al., 2012) has in fact failed to do so. The authors held that, if music did undergo sexual selection, then we might expect to see heightened preferences for more complex music when women are more fertile. In order to investigate this, Charlton et al. (2012) created musical stimuli intended to reflect differing levels of complexity. This was achieved by introducing differing note durations to musical stimuli, and adding syncopation, on a measure devised by Longuet-Higgins and Lee (1984) and later adapted by Fitch and Rosenfeld (2007). An initial experiment was run to confirm that women did perceive the rhythms as differing in complexity, whilst a second experiment asked women to rate how much they liked each musical stimulus. The first experiment found that women did perceive a difference in complexity, but that was primarily associated with a change from isochronous to non-isochronous sequences. The second experiment was the main crux of the paper, and it was found that, whilst women preferred more complex music, no interaction existed between reproductive stage and complexity ratings. In other words, women did not prefer more complex musical sequences around ovulation. This finding disputes the sexual selection hypothesis for modern music. Charlton et al (2012, 5) do state that their findings “do not rule out the possibility that ancestral women used the ability of performers to produce complex music as a criteria for mate choice”.

### 1.1.9 Fukui

One recent hypothesis for the evolution of music alone comes from Fukui (2001), and may serve to cast doubt on Miller’s theory. Having found in a prior study (Fukui, 1998) that music lowers testosterone in males, Fukui (2001) sought to investigate the effect of music on testosterone levels in both males and females. Participants were exposed to music for 30 minutes. There were six musical types, namely the participant’s favourite music, Gregorian chant, Mozart, jazz, pop, and silence. Saliva was collected from participants before and after the study, and testosterone levels extracted.

Overall, it was found that testosterone decreased under all musical conditions, but not to a significant level. However, when the sexes were analysed separately, it was found that testosterone *decreased* in males under all musical conditions (excluding silence). In contrast, testosterone *increased* in females under all musical conditions, again excluding silence. This finding has since been replicated as part of a larger experiment by Fukui and Yamashita (2003).

Fukui discusses the pressures which would have emerged with group living in early humans

- competition for resources of reproduction, including food and mates. This would have led to sexual, aggressive and dominant behaviour due to excessive testosterone. He provides supporting evidence that testosterone is related to libido, activity level, sensation seeking, dominance (Wallen and Lovejoy, 1993), and aggressiveness (Carter, 1992) to frame his theory of an evolutionary role for music. As testosterone responses to libido differ according to gender, Fukui contends that males with lowered testosterone would have had a low libido (Alexander et al., 1997), whereas females with higher testosterone levels would have had higher libidinal capacity (Kemper, 1990), but lower rates of intercourse (Fukui, 2001, 450). He summarises that "decreased testosterone in males would avoid confrontation, and increased testosterone levels in females suppress sexual behaviour" (Fukui, 2001, 450). Thus, Fukui (2001, 450) proposes that music evolved as a system for controlling this sexual, aggressive and dominant behaviour.

This proposal may have an interesting impact on Miller's concept that music evolved through sexual selection. Fukui's experimental test and entailing hypothesis seems to run contrary to Miller's. Rather than music emerging as a *result* of human sexual behaviour, the situation may be reversed - music helped to mediate sexual and social competition. This idea would seem to explain music's spontaneous practice and production "despite energetic costs and lack of survival utility" (Miller, 2000, 337). In fact, there may have been a real, tangible survival utility in music making - mediating competition for resources. This lack of competition could have led to the formation of stronger, more cohesive social groups, which in turn would have led to better use of resources, and thus, increased chances of reproduction. This might provide evidence to contradict Darwin, when he wrote

> "As neither the enjoyment nor the capacity of producing musical notes are faculties of the least use to man in reference to his daily habits of life, they must be ranked amongst the most mysterious with which he is endowed" (Darwin, 1871a, 63).

### 1.1.10 OKANOYA

Kazuo Okanoya (2002) has proposed a musical protolanguage origin for language, based on the elaboration of song and dance used in sexual display. He states that the self-domestication that occurred when our ancestors adapted to a savannah environment freed humans from certain predation risks. This permitted the elaboration of sexual displays, such as song and dance, analogous to the Bengalese finch, whose males have a complex song that they use for sexual display. Okanoya has proven that "complex syntax effectively stimulated the reproductive system of females…complex song patterning should be more attractive to female birds, and therefore, the song syntax in Bengalese finches may have evolved through the process of sexual selection" (Okanoya, 2002, 54) . This is in contrast to its wild relative, the White-backed Munia, whose song is relatively simple. This phenomenon will be discussed in more detail later in this thesis.

Okanoya suggests that this mutual sexual display (song and dance) could serve as a preadaptation for syntax, stating that

> "Since the ability to dance and sing is an honest indicator of the performer's sexual proficiency, and singing is more effective than dancing for broadcasting, singing evolved through sexual selection until it obtained a finite-state syntax" (Okanoya, 2002, 60).

Alongside this development of syntax, he states that semantic calls could develop through the replacement of non-meaningful acoustic tokens used in singing by acoustic calls with a particular semantic content. Thus, a string of semantic tokens could take the role of a public address, leading to the beginnings of a syntactic language. He then states that factors such as neuroanatomical topography and social calculus would have then modified this simple finite-state language into true language.

### 1.1.11 Dunbar

Robin Dunbar, noted for his work on necortex volume and group size in primates (Dunbar, 1992, 2003; Aiello and Dunbar, 1993) has used his expertise in this area to inform debate on the evolution of language. Based on analysis of grooming time, group size and neocortex volume in primates (Aiello and Dunbar, 1993), the authors concluded that language may have evolved at an earlier date than expected, around 250,000 years ago. However, Dunbar now proposes that a date of 500,000 years may be more realistic (Dunbar, 2004; Barrett et al., 2002). This leaves a substantial period of time between the point at which hominin grooming time requirements would have exceeded the 20% limit seen in primates and the emergence of language. Aiello and Dunbar (1993) found a correlation between group size, neocortex volume and grooming time. With larger social groups comes increased grooming time pressures. Extrapolating from the

"Dunbar Number", expected grooming time required would far exceed that seen in primates, and further, would have been unmanageable, with other time demands such as hunting also playing a role. Dunbar states that anatomical evidence does not suggest any kind of step-wise transition from a pre-linguistic stage to a post-linguistic one. In contrast, group size and thus grooming time requirements are hypothesised to have risen "steadily but inexorably on an exponential trend" (Dunbar, 2004, 262). This raises the question of what might have filled the gap between grooming time available and bonding time required. Aiello and Dunbar (1993) proposed vocal exchanges analogous to the contact calls of Old World monkeys and apes. These calls allow bonding to carry on, without taking time away from other essential activities. Aiello and Dunbar (1993) suggest that once the maximum grooming time had been passed, increasing amounts of vocal grooming would have been required to deal with increased group size. Past a crucial point of 30% grooming time spent, language and speech would have evolved, initially limited to the exchange of social information. Dunbar (2004) proposes a shift around 50,000 years ago, from this simple social exchange to exchanges about more extensive cultural matters, such as philosophical questions, religion, and so on. This system of contact calling, Dunbar states, developed into musical chorusing at "quite an early stage in hominid evolutionary history" (Dunbar, 2004, 263), certainly by one million years ago, when Homo erectus group size would have required an investment of 25% of daytime to be invested in social bonding. Dunbar concludes that language is a culmination of a process of increasingly diversified social bonding mechanisms, based on natural communication forms. Language built upon and developed anatomical and neural mechanisms that supported vocal singing, which was initially used as a form of vocal grooming in evolving hominin groups.

### 1.1.12 MITHEN

Mithen (2005) has proposed a musical protolanguage theory which he names "Hmmmmm". This stands for *Holistic, manipulative, multi-modal, musical* and *mimetic*. This system was Holistic - not segmented or combinatorial, Manipulative - used to influence the emotions or behaviours of others (similar to primate calls), Multi-modal - used both vocalizations and gesture, Musical - temporally controlled, rhythmic, and melodic, and Mimetic - utilizing sound symbolism, onomatopoeia, vocal imitation and gesture (Mithen, 2005, 172). Mithen claims that this system was used by Neanderthals as a form of communication.

Mithen, like Okanoya, states that singing and dancing are ideal candidates for the display of good genes, and so, they provided an ideal vehicle for this holistic system. He suggests that

> "We should, therefore, think of the early hominid social vocalizations as elaborations of the type of rhythmic and melodic utterances used by Gelada babboons. If such vocal behaviour were present, we can readily imagine pairs or small groups of hominids that would sound as if they were singing together, rather than gossiping, as a means of consolidating their social commitment to one another" (Mithen, 2005, 136).

This system of vocalizations, gestures and body language took the form of holistic utterances, which functioned as complete messages in themselves. Crucially, he writes that

> "It was only with the emergence of modern humans in Africa after 200,000 years ago, that compositional language evolved, this ultimately being responsible for the appearance of visual symbols, global dispersal, and extinction of other types of human by competitive exclusion" (Mithen, 2012, 107).

Mithen states that, with the advent of Homo sapiens c. 200,000 years ago, segmentation of these holistic phrases could have taken place in a similar fashion to Alison Wray's (1998; 2000; 2002c; 2002b) description of segmentation. Wray states that segmentation could come about when learners mistakenly notice an apparent regularity between two holistic proto-linguistic strings (e.g. if tebima means "give that to her" and kumapi means "share this with her", then the learner might presume that "ma" refers to "her"). Thus, (mistaken) segmentation could operate on a holistic protolanguage until an arbitrary compositional language is reached. Mithen agrees with Wray's segmentation idea, but offers an alternative manner in which segmentation could occur. He suggests that onomatopoeic words, vocal imitation and sound synaesthesia already present in the holistic protolanguage could have created non-arbitrary associations between phonetic segments of holistic utterances and things in the world. These associations would increase the likelihood of the phonetic segments coming to refer to these entities in the world, and thus, perhaps alongside Wray's form of segmentation, compositional language could emerge.

As well as his own discussion of groups of Neanderthals consolidating their social commitment through song and dance, Mithen comprehensively analyses both fossil evidence and comparative studies, and concludes that sexual selection is “more likely to be correct than it may have initially seemed” (Mithen, 2005, 187). His main argument in favour of this is his own “Sexy Hand-axe” hypothesis, whereby the fossil evidence of perfectly symmetrical, beautifully decorated hand-axes supports the view that display was used to attract members of the opposite sex. Thus, song and dance could well have served to attract members of the opposite sex into mating.

However, I would suggest that sexual selection need not be the only interpretation of these elaborately decorated hand-axes. These could have been created in order to signal fitness to *collaborative* partners, not only sexual partners. The same skills and fitness that Mithen suggests would attract a sexual partner - cognitive, behavioural and physiological traits, environmental knowledge, technical skill, planning, flexibility, persistence (Mithen, 2005, 189) would all be attractive to someone who is seeking a collaborative partner. Thus, sexual selection need not necessarily be called on as a selection mechanism operating on a pre-linguistic system. This view will be revisited and tested in chapters four and five.

### 1.1.13 FITCH

Tecumseh Fitch has, in recent years, been a firm supporter of Darwin’s theory for the evolution of language. He draws on music’s universality, cultural variability and generativity to support his claim that music, in particular non-lyrical song, is an obvious parallel to “bare phonology”. His hypothesis is not an entirely novel one, rather a re-working and modernizing of Darwin’s hypothesis. However, as Fitch points out, Darwin’s hypothesis remains remarkably modern, with new data adding to the support Darwin himself cited.

Fitch’s reworked hypothesis is worth quoting in full;

> “Prosodic protolanguage was a system with phonological generativity, using a small set of elements to build hierarchical structures. These structures were vocally generated, voluntarily controlled and learned, sharing core aspects of speech and song. Furthermore, they were culturally shared and infused with an ill-defined “meaningfulness”, but lacked the atomic, decomposable, propositionally-linked meaning that is the central feature of linguistic semantics. . . In short, prosodic protolanguage possessed phonology, and parts of syntax, but lacked lexical propositional semantics”

(Fitch, 2010, 476).

This prosodic protolanguage was a learned rather than innate, generative vocal system, which consisted of movements of the vocal tract (tongue and lips) to create proto-syllables. By virtue of the extant capacity for categorical perception, these proto-syllables constituted a form of 'syllabic discreteness' (Fitch, 2010, 475). Fitch suggests that the exaggerated intonation contours and repetitiveness of "motherese" may be a good model of the prosodic protolanguage. This view will be further discussed in section 1.2.2 on page 28. With regards to meaning, Fitch rightly observes that music is not devoid of meaning, but is not *referentially* meaningful. As Ian Cross defines it, music can be seen to have "floating intentionality" (Cross, 2003). Music can take on the meaning of the environment in which it is used - as Fitch puts it, "music has a kind of free-floating apparent "meaningfulness" that can attach itself, by force of association, to any type of repeated group activity" (Fitch, 2010, 476). This explains the huge importance of music to many human rituals - weddings, meetings of friends, funerals, for example, and its incredible versatility and context-sensitivity.

Fitch does note one key difference between music and language which may be proposed as evidence against the prosodic protolanguage theory - humans are happy to listen to pieces of music hundreds of times, each time enjoying a new aspect of the music. But language is not so repeatable - we constantly innovate new phrases. However, Fitch presents exceptions to this - formulaic and ritualistic phrases, such as greetings, niceties and stock phrases. These play a key part of our linguistic life - they serve to structure our conversation and provide cues to our interlocutors. Fitch suggests these as fossils of a holistic protolanguage stage, which has largely been preserved in music, but, apart from the instances mentioned above, largely lost in language. This precise point is discussed in great detail by Wray (1998, 2000, 2002c), and will be revisited later in the thesis, in chapter three (section 3.1, page 69).

As Fitch and Darwin have both observed, birdsong provides us with the most abundant comparative data, allowing us to inform our hypotheses. Additionally, more species since Darwin's time have been found to be capable of vocal learning - three distantly related species of birds (parrots, hummingbirds, and songbirds) and six distantly related groups of mammals (humans, bats, cetaceans, elephants, seals and pinnipeds) are now known to vocally learn (Jarvis, 2006; Poole et al., 2005). As Fitch argues, "homology allows the deduction of ancestral states, but analogy allows us to assess evolutionary likelihood and test hypotheses about function" (Fitch, 2010, 495). In other words, the fact that these are distantly related species, and that there is no direct evolutionary relationship between birds and humans, for example, means that homology cannot be called upon here - we cannot use the fact that other species are vocal learners to deduce ancestral states. However, we can use these vocal learners to inform hypotheses about the *function* of vocal learning in ancestral hominins.

In summary, Fitch proposes an updated version of Darwin's 1871 hypothesis, which he terms *prosodic protolanguage*. In this protolanguage, phonology came first, hall-marked by the acquisition of complex vocal learning, with simple sequencing, and a hierarchical phrase structure. This created a manipulative, emotionally-grounded communication system, which was used to influence others. As this protolanguage evolved, Fitch proposes that these holistic meanings were broken down into parts, to which referential meaning was attached. This was the beginning of modern language, which Fitch calls the "genetic fixation of the analytic urge". He writes that

> "As the language of its community grew more analytic, pressure for rapid analytic learning by children became strong. This drove the last spurt to our modern language state. . . this last stage seems most likely to have been driven by kin selection, for the sharing of truthful information among close relatives" (Fitch, 2010, 504).

### 1.1.14 Discussion

Many writers have theorised about the possible evolution of language from music. Others have fleshed out the evolution of one faculty, by reference to the other. More theories have touched only briefly on aspects of language evolution from music, such as Skoyles' (2000a; 2000b) singing origins theory of speech, which states that components such as words and syntax, which are necessary for language, can be independently preceded by those needed for song. However, those components needed for song (breath control for example) cannot be preceded by those needed for speech. In contrast to the perceived view that the vocal tract developed to create the wide range of vocal sounds in human language, Skoyles argues that it could have developed in order to provide humans with a variety of musical sounds. He states that speech is only produced using a small part of the vocal tract, opening up the possibility that song may have preceded speech, as speech could not have evolved without these song-based components in place. He postulates the many functional advantages of song, and its easy ability to mix with speech as evidence that a musical proto-stage could have initially arisen and, with further elaboration, have developed into speech.

Another suggestion is that of "tone-dominated affective utterances" as the basis for both vocal music and protolanguage (Fenk-Olczon and Fenk, 2009, 19). These authors use a similar argument to Skoyles (2000a,b), suggesting that

> "Singing without words. . . is cognitively less demanding than the use of a more or less arbitrary code . . . [and] with respect to tonal modulation, singing is more demanding than speech. For this and some other reasons it is more plausible that singing prepared the vocal tract for speech than the other way around" (Fenk-Olczon and Fenk, 2009, 217).

However, this argument deals more with the physical properties of speech and less so with the cognitive developments necessary for language.

One thing that seems missing from many of the above theories (with the exception of Fitch and Skoyles) is the explicit citation of Darwin's original hypothesis. Indeed, Fenk-Olczon and Fenk (2009) cite Brown's (2000) work repeatedly, with no further discussion of his influences. Darwin's short but succinct theory of the evolutionary precursor of both music and language seems a worthy (but sometimes overlooked) hypothesis on which to base further investigation.

As can be seen when one puts these theories into chronological order, there has been a movement away from Darwin's and Jespersen's view that music was prior to language, to a more nuanced theory of an undifferentiated system, which shared elements of both music and language, but was identical to neither. This is the view that will be taken in this thesis - the LSP was a precursor of both music and language, which shared elements of both, and facilitated the evolution of both. The following section will summarise more detailed mechanisms for language origins, before chapter two defines the hypothesis which will be explored for the remainder of this thesis.

## 1.2 PROPOSED ORIGINS OF A MUSICAL PROTOLANGUAGE

### 1.2.1 RHYTHMIC ORIGINS

There is to date some putative evidence of drumming-like behaviour in apes - chimpanzees, gorillas and bonobos. Arcadi et al. (1998) have written of "buttress drumming" by wild chimpanzees - chimpanzees generate low-frequency sounds by hitting tree buttresses with their hands and feet in "discrete bouts of rapidly delivered beats" which are audible up to one kilometre away (Arcadi et al., 1998, 505). This buttress drumming exhibits regional variation – perhaps indicating social learning (Arcadi et al., 2004). Chimps also hit their chest, the ground, and tree trunks (Arcadi et al., 1998, 505). Apes appear to make use of bimanual percussion alongside vocalizations, and this behaviour is a "common, easily observed behaviour in African great apes" (Fitch, 2005a, 9). This drumming is often seen in the context of male dominance and agonistic displays in chimpanzees, bonobos and gorillas, while bonobos have a variety of different clapping and drumming displays. Chest-beating in gorillas is often incorporated into an aggressive display, alongside vocalisations (Schaller, 1963). However, this behaviour has largely been overlooked in discussions of the origins of language and music, perhaps because of the lack of research available into this behaviour. Fitch (2012) has already suggested that the ability to map a motor pulse onto acoustic motor behaviour was present in our evolutionary lineage before the emergence of humans as a separate lineage, and thus we may tentatively, subject to further research, propose this bimanual percussion as a potential origin of a rhythmic precursor to language.

Fitch suggests that isochrony may have developed as a vehicle to synchronise human interactions, or "a by-product of a more fundamental need for synchronised interactions *between different individuals*" (Fitch, 2012, 85). With the evolution of hominin social systems, a system was needed to coordinate groups - as humans exhibit an instinct to form groups and collaborate, which is not seen in, for example, chimpanzees (Fitch, 2012). Fitch states that human rhythmic behaviour is an expression of a more general "'cooperative urge' - a motivation to share experience, activities, and emotions with others" (Fitch, 2012, 86). The fact that beat entrainment seems to have been selected for, or emerged after, the chimp-human split (about six million years ago) seems to lend support to this hypothesis. Indeed, it is as yet unclear whether apes can entrain their drumming to an external beat. There is some evidence that bonobos may entrain their calling in group environments. In particular, bonobos appear to entrain their ‘staccato hoot’ in precisely timed synchrony to a common beat - typically around 2Hz (Merker et al., 2009). This is the only evidence of entrainment to a beat in non-human primates.

The development of isochrony may be linked to the evolution of bipedalism, if we follow the argument laid out by Mithen (2005), who states that bipedalism required the evolution of mental mechanisms to help maintain rhythmic coordination of groups - “as our ancestors evolved into bipedal humans so, too, would their inherent musical abilities evolve - they got rhythm” (Mithen, 2005, 152). Indeed, there is evidence that isochrony aids group coordination, with Brown stating “musical meter is perhaps the quintessential device for group coordination, one which functions to promote interpersonal entrainment, cooperative movement, and teamwork” (Brown, 2000, 297).

This statement is supported by the myriad of work songs which have historically been used to coordinate group behaviour - there is a long tradition of workers singing rhythmic, isochronous tunes, when at a repetitive, group-based task, such as tilling a field. Anthropological evidence suggests that all agrarian societies tend to make use of work songs (Gioia, 2006), and largely, agricultural work songs are/were rhythmic acapella songs, often used with the aim of increasing productivity while relieving boredom (Peek and Yankah, 2004). There is a tradition of African work songs being accompanied by isochronous drum beats (Peek and Yankah, 2004, 52).

However, we do not see remnants of isochrony in language: as Fitch (2012, 85) writes, “isochronicity is dispensable”, and some music making is abstracted away from isochronous beats (the Javanese Gamelan musical cycles for example). Prosody makes use of rhythm, but not in an isochronous manner. So, isochronicity may not provide a good explanation for the origin of a precursor to language. Following Fitch (2012, 87), “isochrony, rather then [sic] being an adaptation for music, may simply be a culturally discovered aid to synchronisation”. We may conclude that isochrony, perhaps an important element to group synchronisation in a system prior to the learned sequential precursor, became lost as a musical precursor evolved to language.

### 1.2.2 ALTRICIALITY AND THE EVOLUTION OF AFFILIATIVE INTERACTION

A further proposal for the origins of music (and perhaps by extension, language) comes from Dissanayake (2000), who argues that "human music originated in perceptual, behavioural, cognitive and emotional competencies and sensitivities that developed from primate precursors in survival-enhancing affiliative interactions" (Dissanayake, 2000, 389). She suggests mother-child interaction as the origin for this. As infants were born prematurely and were entirely dependent on the caregiver for survival, Dissanayake suggests the "coevolution in infants and mothers of rhythmic, temporally patterned, jointly maintained communicative interactions that produced and sustained positive affect - psychobiological brain states of interest and joy" (Dissanayake, 2000, 390).

In other words, mothers and infants coevolved an interactive system using rhythm and melody to fuel the development of emotional and affiliative bonds. This may be supported by the observation that human infants interact with care-givers in what has been termed "proto-musical" behaviour – involving "production of and response to patterns of sound and action" to "entail temporally-controlled interactions involving synchrony and turn-taking" (Cross, 2003; Trevarthen, 1999). This "proto-musical" behaviour is employed in the modulation and regulation of affective state, and the attainment and control of joint attention. This seems to provide support for the multi-modal interactive system proposed by Dissanayake. However, I would contend that this behaviour is not "proto-musical", rather something like "melodo-rhythmic", in which the child benefits from rhythmic training such as turn-taking, alongside melodic contours designed to appeal to emotional and affiliative bonds.

Dissanayake suggests that the "sequentially organized, multimodally produced and processed signals" (Dissanayake, 2000) used between mother and child were found by evolving human groups to be "emotionally affecting and functionally effective" when used and further elaborated in ceremonial rituals where they again served to attune, synchronise, emotionally conjoin, and enculturate participants. This may be supported by the concept of "floating intentionality" proposed by Cross (1999), who states that the "multi-functionality"of proto-musical behaviours derives from the fact that they may be employed in quite different contexts and mean different things to different users. In other words, the interaction which was previously found only in mother-child interaction could be easily appropriated by the group for synchronization, affiliation and enculturation, due to it not being tied to one fixed behavioural environment.

This use of multimodal signals may have then been developed, culturally codified and emancipated as music, apart from the ceremonial contexts (Dissanayake, 2000). Dissanayake concludes by stating that the resemblances seen in infant-caregiver interaction and ceremonial ritual argue for the existence of "an underlying neural propensity in the human species to respond, cognitively and emotionally, to certain kinds of dynamic temporal patterns produced by other humans in contexts of affiliation" (Dissanayake, 2000, 402). So, multimodal signals, previously used to soothe infants when separate from their caregiver, could have taken on a new role, mediating interactions between individuals in a group. This is supported by Mithen's proposal that "music-making is a cheap and easy form of interaction that can demonstrate a willingness to cooperate and hence may promote future cooperation when there are substantial gains to be made" (Mithen, 2005, 214).

This proposal from Dissanayake is well-defended and convincingly written. However, I would qualify Dissanayake's claim by suggesting that a system of group enculturation and synchronization could be further appropriated to signal *fitness*. Dissanayake's suggestion that these multimodal signals could be very effective in attuning, synchronizing, and enculturating group members is further supported by Mithen's claim for their promotion of cooperation. This is exactly why they would be effective as a signal of fitness, as it requires the investment of time, energy and a shared cultural history to learn.

A related suggestion for the evolution of *language* comes from the altriciality hypothesis (Falk, 2004). Altriciality in mammals refers to a state in which newly-born young are incapable of feeding themselves, are relatively immobile and must be cared for by adults of the species. It relates to the proportion of life during which the offspring remains reliant on the parent. With the advent of bipedalism, hominin pelvises had to narrow to allow upright posture and locomotion. Due to the extra pressure imposed by this new form of movement, hominins evolved so as to counteract this effect, by being born prematurely, and thus extending the time spent in an altricial state. Human offspring differ in many ways from primate offspring. Non-human primate infants are born with a strong grasping mechanism which allows them to cling to the hair on their mothers' bodies. However, hominins lost body hair during the course of evolution. Furthermore, as we have seen, human infants are born in a premature state, entirely reliant on their caregivers for survival. Whilst they do have a grasping mechanism, they cannot support their body weight with their grasping limbs. Thus, hominin mothers were unable to carry their infants with them when scavenging and gathering food.

Analysis of this problem led Dean Falk to her "putting the baby down" hypothesis, which proposes that hominin mothers "adopted new foraging strategies that entailed maternal silencing, reassuring and controlling of the behaviours of physically removed infants" (Falk, 2004, 491). In other words, hominin mothers may have placed their babies down, and foraged near them, staying within visual and auditory distance. This allowed them to maintain a bond, despite being physically separate from their children. Furthermore, Falk suggests that early hominin mothers may have elaborated their vocal calls into "affectively positive, rhythmic melodies" (Falk, 2004, 501), which behaved as a form of rhythmic rocking, in order to assure infants that they were safe, and to lull them into sleep.

Falk suggests that the selection for vocal language occurred after this kind of behaviour (routine affective vocalizations towards infants) become common. Falk states that "as mothers increasingly used prosodic and gestural markings to encourage juveniles to behave and to follow, the meanings of certain utterances (words) became conventionalized" (Falk, 2004, 491). Thus, we see the creation of words out of a melodic, rhythmic precursor. Falk suggests that this creation of words could have fed into the shaping of simple instructive utterances within groups. She suggest that at this stage, protolanguage was "in the process of emerging from the prelinguistic melody" (Falk, 2004, 503). She states however, that language retains aspects of the melodic calls it originated in, found in its prosody and tone of voice. Falk's proposal is an interesting one, but somewhat lacking in detail and supporting evidence.

In particular, in her opening section, Falk claims that "the *worldwide* practice of directing musical speech towards human babies provides a temporary framework or scaffold that, among other functions, facilitates their eventual comprehension and production of speech" (Falk, 2004, 491). She further compounds this claim by writing "Motherese is the medium in which infants *around the world* perceive and eventually process their respective languages. . . the view developed below is that parental prosody is not only an *integral* component for propogating language today, it also formed an important substrate for the natural selection of protolanguage in early Homo" (Falk, 2004, 491, emphasis added). Falk authoritatively refers to Motherese and parental prosody as worldwide, and integral. This seems to follow the tradition in literature on child-directed speech, for example "CDS [child directed speech] is presented as a crucial catalyst in the complex process of L1 acquisition" (Matychuk, 2005, 301), and "cross-cultural study attests that child-directed speech is (essentially) universal" (Bornstein, 2006, 903).

However, there is growing evidence that this is not the case. As summarised by Masataka (2003), "evidence for cross-cultural universality of the production of motherese is less convincing and actually very controversial" (Masataka, 2003, 137). He refers to Ratner and Pye (1984), who studied the fundamental frequency in Guatemalan mothers' speech to children, and found that their admittedly small sample of three mothers did not raise their pitch when addressing children. As high pitch and exaggerated pitch contours are considered a key aspect of motherese or child-directed speech (Blount, 1984), this finding argues against the interpretation of motherese as a universal phenomenon. In fact, Ingram (1995) interprets this finding to indicate that prosodic modifications in child directed speech are the result of conventions which vary culturally. Research by Shute and Wheldall (1989, 1995) further supports the idea that motherese or child directed speech is not a universal, static system. They recorded the pitch and pitch range of British women when reading aloud to, and speaking freely with, their children. Analysis did find an overall, average, increase in pitch and pitch range during both discourse conditions. However, the averages here are masking wide individual range differences. Several participants, in fact, did not have substantially different pitch ranges when interacting with children than when interacting with adults. Some participants actually had *lower* pitch ranges and average pitch when interacting with children rather than adults. Masataka concludes from this that "even within a single cultural-lingustic group, more variability likely exists than has been conventionally assumed" (Masataka, 2003, 138). More evidence contrary to the universality of motherese is presented by Ochs (1988) who has shown that traditional Western Samoan societies don't make the changes to speech patterns typical of so-called "universal" Motherese or Child directed speech, when speaking to children. In fact, children are often cared for by elder children, in sight of the mother. Similarly, Kululi speakers of Papua new Guinea do not make such accommodations to children either (Schieffelin, 1985), and in fact children are not spoken to before they master language. Further research questions whether Motherese is actually effective - Newport et al. (1977) found no relationship between the simplification of the mother's speech and the child's linguistic progress.

The evidence outlined argues against Falk's strict definition of Motherese as "worldwide" and “integral” - as seen above, there may be real variation in the extent to which Motherese is used, if at all. As Nwokah writes, “still relatively little is known about cultural differences in the structure and organisation of care-giver speech” (Nwokah, 1987, 213-4). Falk continues to use the concept of universal Motherese as a basis for her theory. As she says “it is argued that selection for vocal language occurred after early hominin mothers began engaging in routine affective vocalisation towards their infants, a practice that characterizes modern women, but not relatively silent chimpanzee mothers” (Falk, 2004, 492). She goes further, making the bold claim that “this universal practice [Motherese] and its associated ontogenetic unfolding of language acquisition in human infants is genetically driven” (Falk, 2004, 492). However, if child-directed speech or Motherese were found to vary cross-culturally, and thus unlikely to be under genetic control, as some evidence I have presented suggests, this would seriously undermine Falk’s claims.

### 1.2.3 Music for social cohesion

One hypothesis for the evolutionary role of music is that it fostered social bonding and group cohesion. This may be as it “encourages the participants to keep a constant audiovisual representation of the collective intention and shared goal of vocalizing and moving together in time - thereby satisfying the intrinsic human desire to share emotions, experiences and activities with others” (Kirschner and Tomasello, 2010, 354). As the authors write, some universal features of music, such as its ritualized context, periodic pulse, discrete pitches and highly repetitive repertoire, make music predictable and aid group cohesion. Indeed, these are the very things we would expect to see in infant-caregiver interaction. As Dissanayake and Falk propose, rhythmic interactions may have served to coordinate emotions between mother and infant, and this, as proposed above, could have developed into a group cohesion role.

Kirschner and Tomasello (2010, 355) state, “the hypothesis of music as a tool for supporting group cohesion predicts that joint music making ultimately increases pro-social commitment and fosters subsequent cooperation among the performers”. This is supported by their study which tested the prosociality of four-year old children. These children were split into two groups, and exposed, in pairs, to one of two conditions; 1) joint music making, 2) joint story-telling (non-musical condition). The conditions were functionally identical, but the musical condition used distinctly musical features such as periodic pulse, use of discrete pitches and a highly repetitive melodic structure, and a discretisation of time and pitch. After this initial interaction phase, the children, still in pairs, were tested on their willingness to help a partner, and to cooperate on a problem-solving task. The "helping" task tested whether children would put aside their own goals in order to help their partner, whilst the cooperation task consisted of a game which could be played alone or in pairs. Results highlighted the important role of music, with children of both genders helping their partner, and playing cooperatively, *more* after joint music making than joint story-telling. In both tasks, girls were more helpful and cooperative, which perhaps tells us something about the sexual dimorphism in children's play. So, results show that "when performed in a manner akin to that typical in traditional small-scale societies, joint music-making enhances pro-social behaviour in four-year-old children" (Kirschner and Tomasello, 2010, 361). The authors state that one of the functions of early musical behaviours may have been "the maintenance of social bonds and prosocial commitment among the members of individual social groups, ultimately increasing cooperation and pro-social in-group behaviour" (Kirschner and Tomasello, 2010, 361).

Edward Hagen and Gregory Bryant hypothesise that music and dance "may have evolved as a coalition signalling system" (Hagen and Bryant, 2003, 21). They state that rhythm, which plausibly serves to synchronise musical and dance performances by groups, is not found in any non-human primate, and that the sexual selection hypothesis offers no explanation for this unique feature of human music and interaction. This echoes Fitch's suggestion that isochrony developed as a means to synchronise interactions between humans (Fitch, 2006a, 2010). Hagen and Bryant go on to say that group cohesion and intergroup interactions are best served by the exchange of credible information. They hypothesise that groups may have used music and dance to signal both that their group was internally stable, and that it had the ability to execute rapid, complex, coordinated action (Hagen and Bryant, 2003, 29). Music and dance were suited to signal this, as they could be easily decoded by receivers, and they showed credibly that the group was internally stable, was likely to remain internally stable, and had the ability to execute complex coordinated actions. The signal was honest as it was, out of necessity, correlated with the quality it was made to present – it could not be faked. Thus, the time and energy needed to create and practice music corresponded to the time that the group members must have been in close collaboration. This forms the basis of Hagen and Bryant's *Coalition Quality* hypothesis. A group that had been in contact with each other for a short time could only create and faithfully reproduce a relatively simple performance, whilst a group which had been in close coalition for a long time, and thus had close collaborative ties, could perform a vastly complex piece, thus signalling their coalition quality.

This hypothesis has many strong points. I believe we can take a lot from it, in shaping a new theory of the evolution of both music and language. As Hagen and Bryant rightly say, music is a reliable indicator of some qualities. It can be easily decoded by an audience, whilst it is much more difficult to learn and compose, taking time and cognitive capacity to do so. However, I propose that this ability need not have signalled coalition at a *group* level, but cognitive ability at an *individual* level. This point will be discussed further in chapter two. An important point to take from Hagen and Bryant's paper is that "music and dance do not *cause* social cohesion, they *signal* social cohesion" (Hagen and Bryant, 2003, 30). Their hypothesis also appears to explain musical variety -

> "Because membership in human coalitions is relatively fluid, music could not signal coalition quality if there were only one song – everyone in the population would simply learn the one song, and performance of this song would then reveal nothing about the quality of specific coalitions" (Hagen and Bryant, 2003, 31).

Thus, this musical system must have been rich in repertoire.

We may also interpret this claim in light of our own hypothesis, outlined in brief in chapter 1, beginning on page 1. Musical variety may be explained by the need to signal *individual* cognitive ability. Were the musical system used by our ancestors to collapse down to one song, there would be no way to monitor and signal individual fitness, and the system would quickly become unusable and collapse.

A further proposal along these lines was made by Hagen and Hammerstein (2009). They propose a proto-music which lacked several features we see in modern human music. However, with increased human interaction, "the need to signal identity adds two of the missing features of music - learning and variation - to protomusic" (Hagen and Hammerstein, 2009, 299). They further state that

> "Choosing allies based on the quality of music and dance performances . . . would spark an evolutionary arms race between coalition members with an interest in producing ever more convincing signals of coalition quality, and potential allies with an interest in better discriminating between performances of coalitions of different

> quality, leading, eventually, to the rich coalition signalling system we call music" (Hagen and Hammerstein, 2009, 300).

Additionally, Hagen and Hammerstein state that "music and language might have played complimentary roles in the emergence of the sophisticated agent-like properties of human groups ... music might have evolved to credibly signal coalition quality (and perhaps also identity) ... language, in turn, might have evolved to play a complimentary role" (Hagen and Hammerstein, 2009, 301).

As Hagen and Bryant themselves state: "the coalition quality hypothesis does not account for solo performances" (Hagen and Bryant, 2003, 40). So, we may take aspects of their theory and reanalyze it in light of our own.

## 1.3 Discussion

In just under 150 years, the field of musical protolanguage theories, and research into the origins of music and language, has grown rapidly. However, little to no empirical research has explicitly tested these theories. Darwin's initial writings on a musical ancestor of language are concise, clear, and have inspired remarkable scholarship in the years since 1871. Tecumseh Fitch, Stephen Brown, and Stephen Mithen in particular have drawn on aspects of Darwin's theory in creating their own hypothesised musical protolanguages, be they termed "Prosodic protolanguage", "Musilanguage", or "Hmmmmm". The views detailed above largely share some base elements - some music-like or prosodic system which aided in group cohesion, sexual or territorial issues, which underwent selection, and eventually bifurcated into language, in which a pressure for expressivity evolved, and music, in which a pressure for emotion evolved. The two systems now share many similarities, but have at least one key functional difference - language can be referential, while music has no concrete reference, but rather a kind of "floating intentionality", a non-referential meaning, which can be distinct to each listener.

Many different views of the function and origins of these protolanguages have been put forward - infant-caregiver interaction, group interaction and cohesion, mediation of testosterone and thus competition. It should be noted that these evolutionary theories need not be mutually exclusive. The concept of some kind of music-like system acting as a social bonding system is interesting, and will be discussed as a previous instantiation of a musical signalling system in the following chapter. The theories outlined above have inspired this thesis, in particular the clear need for empirical investigation. Chapter two will outline the hypothesis presented in this thesis, chapter three will root it in time, while chapters four and five provide empirical support for this particular hypothesis, and suggest the beginnings of a rigorous empirical approach into the origins of language in music.

# CHAPTER 2

# A NOVEL HYPOTHESIS FOR A MUSICAL PRECURSOR TO LANGUAGE

In this chapter I will present a hypothetical system of signalling cognitive abilities through complex, sequential, culturally learned signals, in particular melodic and rhythmic signals. Rhythm and melody were both key to this system, and remnants of this Learned Sequential Precursor can be seen in both music and language, which make use of both rhythm and tone to varying degrees. This precursor is seen as having come *before* a protolanguage, which may or may not have retained the musical form of the precursor. It is outside of the scope of this thesis to prove or disprove the exact origin system or systems of a learned sequential precursor. Some possible precursors (altriciality, group cohesion) are proposed in chapter one. However, it is with the next step, the proposed appropriation of this original system to signal cognitive abilities, that the thesis begins. In this chapter I will propose a system that signalled cognitive abilities, which may have undergone social selection as a cue to inform collaborative partner choice. This complex learned musical precursor may, with cognitive and social evolution, have been adapted into meaningful linguistic and emotional musical systems. The relevance of theories of birdsong, developmental stress and stability, identity and domestication will be discussed, before likely steps to language and music are outlined.

## 2.1 A NOVEL HYPOTHESIS - THE LEARNED SEQUENTIAL PRECURSOR (LSP)

Music is a wonderful signalling tool. We use it to signal to others our emotions. As we have seen above (section 1.2), there are various proposed origins for a musical protolanguage. These include primate vocalisations, primate drumming (Fitch, 2006a), and infant-caregiver interaction (Falk, 2004; Dissanayake, 2000), among others. The hypothesis to be proposed in this section is also seen as a precursor to protolanguage, crucially as a signalling system. We can propose how this original system may have enabled the emergence of a learned sequential precursor to protolanguage. As hominin groups developed, through thousands of years of evolution, group interactions became increasingly important to hominins, enabling cohesive hunting, gathering and shelter-building. In group interactions, let us presume, some group coordination system may have been in place. This could have taken the form of a group cohesion system as discussed by Dissanayake (2000) or Hagen and Bryant (2003). This would have been a ritualised system with no referential meaning, which was available to all group members. However, in order to exploit the skills of group members, and thus resources, adequately, it would have been advantageous for early hominins to signal their abilities and skills, in order to engage in collaborative interactions. This hypothesis will focus on *cognitive ability*.

A group that interacted using a ritualized multimodal, rhythmic system would have created a strong basis for group trust and cooperation. It is proposed that a sub-set of the population may have become able to further use this coordination system for their own means, which would have given them a large advantage over others. This subset of the population would have appropriated the multimodal system in order to signal their own cognitive ability, by adapting the ritualized, habitual coordination system into a complex musical system which required an investment in learning, innovating and producing, the LSP. See Figure 1.1 on page 3 for a graphical representation of this change.

The ability to learn to use this musical precursor, which may have depended upon cognitive ability, alongside a requisite investment in learning, and crucially, membership of a cultural group, acted to keep the signalling system honest, as only those who could spare the cognitive capacity and time (which could be spent learning from the group or hunting, for example) could learn and reliably reproduce a complex sequence of actions. The requisite cognitive abilities, in fact, mean that the LSP was an index to the quality of the signaller. We might use the Investment in Learning Hypothesis coined by Kirby (2012) to describe the learning of this complex system. Under this hypothesis, the system learnt must take an investment - in this case, cognitive capacity, physical energy, and time. This system is referred to as a Learned Sequential Precursor, meaning that complex learned sequences of multimodal sounds were in use. We can borrow from Fitch's (2005b, 132) discussion of human evolution, in which he states that "the crucial first step in human evolution was the development of vocal imitation…This augmented the already-present movement display behaviour seen in modern chimpanzees and gorillas to form a novel, learned, multimodal display system". Crucially here, the LSP was a novel, learned, multimodal display (signalling) system.

This complex signalling system, subject to cultural transmission, must have been difficult to learn, but not difficult to perceive. It must have been complex enough to disallow cheating, but simple enough that variation in quality could be easily perceived. Also, the system must have been beyond what can be encoded in genes (although it was built upon and facilitated by genetic changes such as encephalisation). Thus, the system was complex but crucially, changeable, in order to evade cheaters. Cheaters here are seen as those who could potentially learn the system without a major investment, thus benefitting from the potential collaborative act, without having invested in the group.

As discussed earlier, the LSP is proposed as an honest signal of cognitive ability. This could be used by perceivers as a signal of the producer's quality as a partner. However, I do not propose this as a sexually selected trait, but rather one which informed *collaborative* partner choice. Being chosen as a collaborative partner may have been correlated with an increase in fitness, as successful signallers who collaborated would benefit from the interaction, and thus live longer and lavish more resources on their offspring, leading to increased fitness. As fitness-increasing traits increase in frequency through selection (if genetically underpinned/heritable), natural selection should increase the frequency of the trait which improves the likelihood of being chosen as a collaborative partner. This is proposed to be music-relevant cognitive abilities, which are proposed to have a genetic basis, and to be a reliable indicator of cognitive ability in other domains. Thus, we might expect to see these genes being selected for and increasing in the population, assuming that social selection driven by collaborative partner choice was informed by the musical output that results from these genes.

The following diagram shows the relationships proposed here. Musical ability increases the chances of being chosen as a collaborative partner. This leads to increased resources for both partners, and thus increased fitness. So, we should expect to find genes which underpin music-relevant cognitive abilities (as the trait which improves likelihood of collaborative choice) to reappear in increased numbers in subsequent generations.

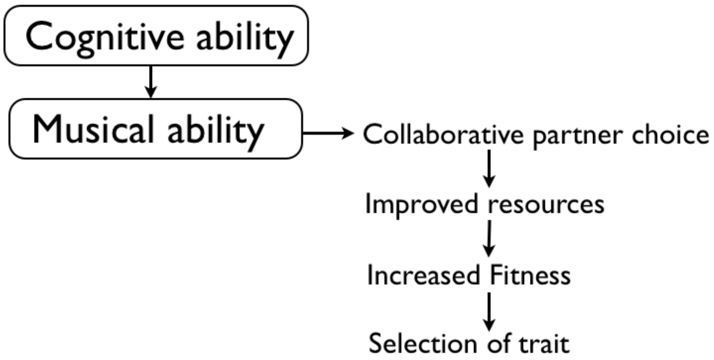


Figure 2.1: Relationship between cognitive ability and musical ability, showing the operations of collaborative partner choice and selection

It is prudent at this point to discuss the choosers in this system - as mentioned above, selection of the LSP is proposed to have been driven by collaborative partner choice, as perceivers used the musical output of the signallers as a cue to their cognitive ability. But *why* were perceivers choosing? This is a challenging question. We may address this in part by reference to the relative percentage of high-quality signallers in the population (i.e. those with music-relevant cognitive abilities). This is proposed, initially, to have been low. So, we have a small minority of the population who signalled *high* cognitive abilities. Let's imagine one perceiver of the musical signals then collaborates with a high quality signaller. As a result of the signaller's abilities during the collaborative task, both signaller and perceiver get increased resources. This should lead to selection, as a drive to collaborate with the still relatively rare high quality signallers, in order to increase one's own fitness. So, we should find that the propensity to collaborate with high-quality signallers would increase in frequency due to these fitness-increasing effects.

### 2.1.1 Manual skill, cultural learning, and creativity

This learned sequential precursor was a complex system, which required many interweaving levels of learning, production and creativity. We may think of this system as an onion, with many layers. The centre of the onion is motor skill. This is key to being able to use the precursor. Let's imagine one clever hominin, who realises that he can create sound by hitting his chest with his hands, banging on a hollow tree trunk, or clapping his hands. This exhibits good motor skill, and also some flash of inspiration. However, this new skill does not demonstrate ability to learn, rather simply motor ability. So we come to the next layer of the onion, cultural learning. This level requires that individuals learn a cultural system, and replicate it faithfully. The individual must have invested in a cultural group in order to learn the musical culture in place. At this stage, however, this is made up of a set of culturally-appropriate sequences (rhythmic/musical pieces, or in the case of our clever hominin, a series of claps and bangs), which are copied with little to no alteration. So, we have a simple copying system, which most of the population are adept at interpreting - both producers and perceivers must have invested in this cultural system.

It is when we reach the outer layer of the onion that the complexity and expertise of the system comes into play. This layer requires the ability to reproduce a cultural system, adding creativity as an important display mechanism. Creativity involves having both motor skill, and a share in the cultural system. The creative individual learns the cultural system, synthesises the information from different sources (here different rhythms/ songs), and takes this one step further, *by creating his own rhythms*. Thus, a musical system could come into being. All three layers of the onion must be in place for the system to work - creativity in production doesn't exist in a vaccuum, and so relies on cultural learning, which in turn relies on the core motor skill, for its existence. Producers must have mastered all three stages - motor skill, cultural adherence and creative production. However, perceivers need not be able to produce signals in order to judge the musical output of producers. Both producers and perceivers must be aware that information is potentially held within signals, and able to derive this information. This is analogous to birdsong, in which (generally) males sing and females perceive. Even though they do not display, females must be able to judge the quality of the male's song, and so, they are users of the system too. This interplay of three crucial steps created a system of differential signalling of cognitive abilities, which required cognitive ability as well as an investment in learning and the cultural group from producers.

The interplay of these three crucial systems justifies the investigation of music as a precursor to language. All three elements may be seen separately in different systems; for example, motor skill can be seen in many manual jobs, such as bricklaying, and cultural learning is important to many systems and roles. Other examples of systems which use interplaying roles of motor skill, cultural learning and creativity may be painting or artwork. The LSP need not necessarily have been musical in nature. This hypothesis equally supports other culturally learned systems, such as dance, and displays such as Mithen's "Sexy Hand-axes" (Mithen, 2005). Any activity which required the learning of sequential signals would have been facilitated by this development. However, I chose to investigate music due to the relative ease of use of auditory material in experimental research and the growing field of research into this topic.

### 2.1.2 Properties of the Learned Sequential Precursor

Rhythm can be seen as a complex sequential behaviour, and melody can be intensely complex. Stephen Mithen has written that "music and dance are ideal candidates for display" (Mithen, 2005). The reasoning behind this claim is intuitive - singing consists of raising the voice and broadcasting over a relatively wide area, while dancing can attract observers, and also serves as a signal of symmetry, physical strength and fitness. Thus, the learned sequential precursor here proposed as a precursor to protolanguage consisted of multimodal song and possibly dance. The vocal system made use of the evolved hominin vocal apparatus and articulators creating what we might call "proto-syllables", following Fitch (2010). This could have been a very early substantiation of phonology. No compositional syntax is proposed at this stage, and there is no referential meaning to be considered. This precursor to protolanguage is proposed to consist solely of an honest system for the signalling of cognitive ability.

As mentioned in the introduction to this thesis, this chapter outlines a learned sequential precursor, determined as "music-like" in nature. This term is used, as it would be misleading to assume that the precursor to language was similar or identical to music as we know it now. Rather, it was likely quite spartan compared to the rich music we listen to today. Fitch (2006a) has devised a number of design features of music, inspired by Hockett's (1960) design features of language. Under this analysis, music; is complex, generative, culturally transmitted, consists of discrete pitch, is isochronous and transposable, exists in a performative context, is repeatable, and is a-refererential. Here, Fitch (2006a, 176) means by vocal music “all music generated by the vocal tract but lacking distinct words”. With reference to the LSP, it is difficult to determine which of these features may have been in place. It seems reasonable to assume that the features of cultural transmission, generativity, complexity and a-referentiality would be found in the LSP. The LSP is determined to have been generative as it consisted of discrete pitches which could be recombined, although it was a-referential. The LSP was complex, as will be investigated in chapter four. It was culturally transmitted through its use as a cultural system.

The key distinguishing factor between music and the LSP may be that of performative context. It would be hasty to claim that this LSP was only used in distinct, determined contexts - rather, it was a signalling system which may have been used in many contexts. Other features may be plausible, but impossible to adequately determine. So, we must conclude, based on the deviations from Fitch’s (2006a) analysis, that this LSP was not music, but “music-like”. Throughout this thesis, the term “musical” when used in reference to the LSP, should be interpreted as meaning “music-like”, as defined here.

### 2.1.3 Rhythm

The proposed precursor to language is a rhythmic one. Thus, we must analyse rhythm as a human skill. Although, as Fitch (2010) states, periodicity and entrainment are ubiquitous to living organisms, humans are the only living beings who entrain to an external beat. Fitch recognises that what we consider “human rhythmic behaviour” can be split into (at least) three elements, each with their own biological basis and evolutionary history. The first of these is “periodic motor pattern generation”, which Fitch states is ancient and ubiquitous. This is the ability to generate a regular beat. Although Fitch does not address this, this may be dependent upon bipedal locomotion. Mithen (2005, 2012) has discussed the likely “immense” effects that bipedalism may have had on human musical capacities - “as our ancestors evolved into bipedal humans, so too would their inherent rhythmical abilities evolve” (Mithen, 2012, 106). Evidence that “the tempo range of optimal musical pulse perception, around 300-900 ms onset-to-onset, is similar to that of locomotion” (Trainor, 2007, 17) seems to support this claim.

The second element of human rhythmic behaviour is ““pulse” (or “beat”) extraction from complex patterns” (Fitch, 2012, 77). Fitch claims this is a form of perceptual cognition that is shared with speech. However, Fitch claims the most unusual feature of human rhythmic behaviour is the third element, “entrainment of one’s own motor output to this inferred ‘beat’”. This is what makes humans dance. As Fitch (2012, 77) writes, “one does not, consciously, go through separate steps of pulse extraction, beat synchronization, and pattern generation” - rhythm is a composite system, which contains both primitive, biological elements, and highly unusual, recently evolved, elements. Fitch states that periodicity and entrainment are “among the most basic features of living things” (Fitch, 2012, 78). Synchronisation often occurs between coupled nonlinear oscillators (Bennett et al., 2002), and Fitch (2012) assumes that humans exploit this synchonisation in rhythmic entrainment. The “paradox” that Fitch discusses is not that human do exploit this, but that few animals appear to. Exceptions identified thus far are fireflies, some amphibians, and parrots (Greenfield, 1994; Merker, 2000; Patel et al., 2009; Schachner et al., 2009).

Fitch uses this lack of synchronisation (with few exceptions) to support the assertion that “beat entrainment either emerged, or underwent positive selection, sometime after the split between chimpanzees and humans around six million years ago” (Fitch, 2012, 81). Fitch further addresses isochrony. As seen in chapter one, Fitch proposes that isochrony developed as a vehicle to synchronise groups, driven by a desire for “group coherence and predictability” (Fitch, 2012). He states that isochrony results from an interaction of the pattern extraction and entrainment aspects of the rhythm faculty, as discussed above.

Recent work (Schachner et al., 2009) has lent support to the claim from Patel and Iversen (2006) and Patel et al. (2008) that the capacity for entrainment evolved as a by-product of selection for vocal mimicry. Schachner et al. (2009) analysed thousands of online videos from YouTube, a video sharing website, which featured animals “dancing”. They found that only vocal mimicking species showed signs of beat entrainment, according to their two measures (maintaining consistent phase and matching frequency). Of these vocal mimickers, 15 species showed “evidence suggestive of entrainment” (Schachner et al., 2009, 834), these species including 14 species of parrot and one elephant species. Vocal non-mimickers showed no evidence of entrainment. This supports the claim by Patel and Iversen (2006) and Patel et al. (2008) that selection for vocal mimicry is a necessary precondition for entrainment, and that this selection modified the basal ganglia, creating an auditory-motor coupling which forms the basis for beat entrainment. These data, while interesting and perhaps inspiring of other research, fall foul of the same criticisms which could be levelled at the Patel and Iversen (2006) study - this study is based on videos created by animal owners, and may not be as empirically rigorous as would be preferred.

## 2.2 LEARNING

Key to the LSP hypothesis is the learned element of the system. The proposed sequential precursor is a learned skill: both signaller and perceiver must learn the system before they can engage in a successful and appropriate signal-response interaction. It should not be overlooked that both potential interactors must be members of the same cultural group for this interaction to be successful. However, whilst it is relatively easy for the perceiver to learn to perceive quality in a signal, it is very difficult to learn to produce such quality.

To signal their abilities, the signaller must invest time and energy in the cultural group (as does the perceiver to a lesser degree). The signaller must learn the culture-appropriate signals and responses. However, he is constrained by his own cognitive abilities. He cannot fake a high quality signal if he is not himself of high quality - as Bolhuis et al. (2010, 748) state: “what can be learned is constrained in both species [birds and humans] by morphology and physiology”. We can follow from Kirby’s (2012, 98) discussion of the *Investment in Learning Hypothesis*, in which he states that fitness indicators only work “in the particular case where the song being transmitted culturally is in some way hard to learn - in other words, where acquisition is challenging”. Here he is using the case of birdsong, but this could equally well apply to our proposed learned sequential precursor. Our hypothesis is that of a culturally transmitted system which is complex enough to take non-trivial investment to acquire, and requires group membership. Thus, differing ability levels could reflect (cognitive) capacity to learn this skill, as well as level of commitment to, and investment in, the group. This makes the signal, when produced and perceived correctly, an honest, accurate signal of ability to learn the LSP, which is underpinned by music-relevant cognitive abilities. As Kirby (2012, 98) writes, “in forming coalitions for joint activity, accurate assessment of other’s abilities may be invaluable”, and so, the reliable signal of another’s ability supplied by this learned sequential precursor would be useful.

Kirby (2012, 98) outlines what we might expect a rhythmic structure to look like if it were used as a reliable signal. It should be "hard to learn to produce. . . it need not be hard to learn to perceive. . . [and] it should exhibit some degree of cultural variation". This neatly sums up our proposed learned sequential precursor, which has an asymmetry of production/perception and is culturally learned and transmitted.

We may draw some strong analogies between this learned sequential precursor and birdsong. As Bolhuis et al. (2010, 748) state, in both humans and birds, "learning is the product of the interaction of predispositions and specific experiences. . . however, what can be learned is constrained in both species by morphology and physiology". Both male and female birds must learn the song system of their species, which entails energy expenditure. However, the male must signal his fitness as a sexual partner through producing a song designed to impress the female. This song acts as a reliable signal, as degree of resistance to developmental stress (in the shape of early life lack of nutrition or parasites), is communicated in the song of a male bird (according to the Developmental Stress Hypothesis of Nowicki 2002 and Spencer 2003). This hypothesis argues that song learning can be seen as an indicator mechanism which allows females to use learned features of song as a window on a male's early development, which will highlight any developmental stresses which may have long-term phenotypic effects (Ritchie et al., 2008, 570). Whilst the example of birdsong differs in function from the LSP, which is proposed to have signalled cognitive abilities, I feel that birdsong, and some experimental research related to birdsong, is still a rich and relevant are from which to draw inspiration.

As Ritchie and Kirby (2006, 283) state, "the evolution of learning can. . . be seen as a key transition in the evolution of human language". In order to investigate possible ecological factors affecting the transition to a learned communication system, Ritchie and Kirby created a computational model of bird song development. Bird song is useful in this investigation, as it has striking developmental similarities to language. Nestlings and babies both have a critical period for learning, both rely on auditory feedback to develop normally, and both exhibit a form of babbling or subsong (Doupe and Kuhl, 1999).

Catchpole and Slater (1995) identify an auditory template of song development in birdsong. This consists of two phases, the memorization phase (which all birds experience – we might profitably label this the perception learning phase), and the motor phase (which only the male birds experience - which we may label the production learning phase). In the memorization/perception learning phase, during infancy, an innate crude template is modified by the hearing of the song of the bird's own species, to create an exact template. This identifies the song behaviour of perceivers, the female birds. The males continue to the motor/ production learning phase, in which the bird's own song is matched to the exact template created during the memorization/ perception learning phase. Thus song production is trained to produce songs resembling the learned template.

Ritchie and Kirby took this model as inspiration for a computational model of the two stages of learning in birdsong, namely observational learning and reinforcement learning. They determined an agent's fitness by its ability to recognize and be recognized by conspecifics. Having this measurement of fitness meant that there was a strong selection pressure for agents to develop and maintain an easily recognised, stereotyped species-specific song (Ritchie and Kirby, 2006). Ritchie and Kirby manipulated only two simple ecological factors: one being environmental reliability, i.e. the degree to which previous generations' songs are faithfully recorded and passed on to a new generation to learn from, the second being timing of song requirement, i.e. whether song is required immediately from birth. They created two environments, one an unreliable environment where only 20% of songs are passed on, another a reliable environment where 80% of songs are passed on. They found that in reliable environments, genes need not code for song, as agents can rely upon cultural transmission to get a copy of the correct song. Thus, learning can be relied upon to convey song, meaning that genes need not do so. Additionally, the simulated agents were seen to rely only on their auditory copy of their song, rather than storing both a sensory and a motor copy. Thus, the male bird only produces a song when it is sexually mature, while both male and female birds need to be able to recognise song before this point. Ritchie and Kirby (2006: 290) state that "the song recognition system should be more genetically constrained than the song production system". This supports our hypothesis that perception should be easier than production, in an environment where everyone has adequate knowledge of the system. As discussed in section 2.1.1, all individuals who interacted using this learned cultural system (LSP) must be able to judge the quality of producers, whilst only a subset of the population attained the production ability (due to an investment in learning) necessary to signal musically. The following sections will address mechanisms maintaining the honesty of the LSP.

## 2.3 DEVELOPMENTAL STRESS

Recent empirical work with male zebra finches (Boogert et al., 2008), has presented the first evidence of a link between a bird's song complexity and its speed at learning. Stating that "females selecting males on the basis of song complexity may, in effect, be selecting for particular neuro-anatomical qualities that underlie both song development and cognitive processing in general" (Boogert et al., 2008, 1736), the authors investigated the link between song complexity and performance on a novel foraging task. They recorded the directed songs of 27 male zebra finches, by recording the songs when in the presence of a female zebra finch. These directed songs are hallmarked by faster tempo, less variable syllables, and the fact that birds are less likely to begin singing in the middle of a phrase when courting females than when singing alone (Boogert et al., 2008, 1737). The recorded songs were quantified for complexity by measuring three aspects - average song phase duration, total number of elements, and number of unique elements, per song phase. Then, each male was presented with a task which required him to forage for seeds in four levels of difficulty. This foraging task required the birds to gain access to seeds in a food well. In the first level, the lids were positioned next to the wells, in the second level, half of each well was covered by the lid, in the third, the wells were fully covered by the lids, and in the fourth, final level of difficulty, the lids were firmly fitted into the wells. In order to master the final level of difficulty, the birds were required to learn and remember all skills necessary to gain access to the food source, in prior rounds.

Results showed that one measure of complexity - total number of song phase elements - was a significant predictor of learning performance, with males whose song phases contained more elements requiring fewer trials to solve the final level in the foraging task (Boogert et al., 2008, 1738-9). This supports their "cognitive capacity hypothesis", that song complexity might be an honest indicator of learning capacity. In fact, song complexity can be seen as indexical of cognitive abilities relevant to foraging. As the authors write, "the cognitive capacity to learn how to adapt quickly to variable foraging conditions might make the difference between life and death for individuals of many bird species" (Boogert et al., 2008, 1739). This claim could easily be adapted to human abilities - the cognitive capacity to hunt well, plan, and adapt to new situations could have led to differential fitness of signallers, through differential collaborative choice. This fascinating study lends credence to our hypothesis that complex sequences of actions (musical productions) may have served to signal cognitive ability, and indeed the experimental approach taken could be adapted to suit human experimental purposes.

For obvious ethical reasons, we cannot tinker with the developmental stress or stability of human subjects, and so we must draw on birdsong research in order to inform hypotheses of language origins and evolution. Thus, research into birdsong development has been very influential in informing the LSP hypothesis. As Fitch argues, “homology allows the deduction of ancestral states, but analogy allows us to assess evolutionary likelihood and test hypotheses about function” (Fitch, 2010, 495). We can and should use analogous birdsong to inform our investigations of language origins. For example, birdsong is a vocally learned system. The system is honest and unfakeable because it takes an investment of cognitive capacity and physical energy, alongside time, to learn and reproduce. All of these aspects (investment of time, cognitive capacity, and physical energy) are relevant and important to the proposed LSP. This system is proposed as a culturally learned, vocal and gestural/dance-based signalling system, which required an investment of time and energy. This system was kept honest by the cognitive ability and cultural investment required, and by the fact that it must have evolved in socially-living, co-dependent groups of early humans. This discussion of two mechanisms which served to maintain honesty will be expanded below.

## 2.4 GROUP MEMBERSHIP AND HONESTY

As Ritchie (2009, 80) writes,

> for an indicator to function reliably there must be some mechanism enforcing the correlation between the condition of the signaller and the form of the signal. . . That is, for an indicator system to be evolutionarily stable there must be some mechanism ensuring the *honesty* of the system.

This view is important for the evaluation of the LSP. This section will examine differing ways in which honesty may have been maintained in the LSP. First, the role of cultural identity and investment in group membership will be examined, looking at these as examples of handicaps. Secondly, the signal itself will be discussed as an example of an index of music-relevant cognitive ability. Thus, two mechanisms maintained honesty of the LSP. Cultural investment and the related handicaps alone could not account for the reliability of the system. However, the crucial indexical relationship between signal and music-relevant cognitive ability enforced honesty, which was further supported by the costly handicaps of group investment.

Before we further discuss honesty in signalling, it is important to define what is meant by the term "signal". There are two key conceptions of signals in biology. The first simply states that biological signals convey or carry information. The second, from Maynard Smith and Harper (2003, 3) defines a signal as "any act or structure which alters the behaviour of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved". An example of this might be seen in birdsong - birdsong is designed to affect the behaviour of others, and has evolved because of these effects. Furthermore, the response to birdsong has evolved to be affected. For example, birdsong is a clear signal of the developmental history of an individual bird, and so his song will affect the response of a female bird accordingly. A high quality song indicates that the male bird is of high genetic quality, and thus a worthwhile potential partner. On the other hand, a low-quality song indicates the male's low quality as a potential mate. The song system has evolved in order to convey this information. On the perceiver's side, the female bird uses the song as a reliable cue to the genetic quality of the male, and thus may choose to mate with him. The song system has evolved to have this effect on the behaviour of the female, and the female's response has evolved in tandem.

Darwin (1871a), in observing seemingly deleterious traits such as deer antlers (which require energy to grow and sometimes cause the death of fighting deer), peacock tails, and colouration and displays by birds, both of which attract predators, noted that these sexually selected ornaments are "acquired at the expense of greatly impeded powers of flight or running" (Darwin, 1871a, 453). Darwin suggested that the disadvantages associated with these characters must, in order to be stable, be compensated for by increased mate preference for these animals. In other words, the display associated with having antlers, colourful plumage and loud displays, must attract more and better females to mate with these males. However, Darwin was unable to explain why females might be drawn to these displays over other, less showy male displays. It is an observable fact that peacocks with the most impressive plumage are preferred by females, but "there is no simple explanation to suggest in what ways the preferred males should be of better quality than others" (Zahavi, 1975, 205). This observation led Zahavi to his "Handicap principle". This principle states that

> "characters which develop through mate preference confer handicaps on the selected individuals in their survival. These handicaps are of use to the selecting sex

> since they test the quality of the mate. The size of characters selected in this way serve as marks of quality" Zahavi (1975, 205) .

This Handicap principle essentially holds that honesty is maintained by signals being costly, and so only high quality individuals can afford to pay these costs. In relation to the observations made by Darwin, the theory goes that males can signal their fitness through exhibiting un-necessary and potentially harmful traits, for selection by females. Zahavi imagines this handicap as a test of sorts; "an individual with a well developed sexually selected character, is an individual which has survived a test. A female which could discriminate a male possessing a sexually selected character, from one without it, can discriminate between a male which has passed a test and one which has not been tested" (Zahavi, 1975, 207). So, a bird with a colourful plumage has passed the test as he has already survived the potential increased predation associated with his plumage. This signals his quality as a mate to females, who can select him as a partner, thus selecting for quality and increasing the preponderance of this sexually selected trait. The handicap excludes faking, as it requires such an investment and increases risk. This required investment makes cheating unprofitable.

According to Zahavi and Zahavi (1997, 229), "by wasting one proves conclusively that one has enough assets to waste and more". The crucial point here is the "and more" - the assumption being that high-quality signallers should be able to waste money or viability, and yet still have more residual money or viability, resulting in a positive correlation between amount wasted and residual amounts left (Getty, 2002, 364). As (Pinker, 1997, 500) writes, "conspicuous consumption. . . works when. . . not *all* of the wealth. . . is sacrificed. If I have a hundred dollars and you have forty, I can give away fifty, but you can't. I will impress others and *still* be richer than you".

We can see analogies to the musical learned sequential precursor in Zahavi's discussion of Handicaps, and more broadly in costly signalling theory. The ability to signal well required a large investment on the part of the signaller - an investment of cultural time, which may mean less time hunting or finding resources, and energy, which implies that increased food needed to be processed. These, taken together, show that a signaller who was handicapped through learning this system, and yet, had the ability to signal well, and had survived despite this handicap, had resources to draw on, and thus was a useful partner for collaborative tasks. This musical learned sequential precursor required an investment in learning, as part of a cultural group, which conferred a handicap on the user. This was one mechanism which supported the honesty of the LSP. As Zahavi (1975, 207) writes, “the more developed the character the more severe was the test” - the better the signaller, the more it has invested in learning.

As we have seen above (section 2.2, page 45), both the producers and perceivers in our early human population must have socially learned the LSP, in order for a stable system of signalling abilities to evolve. Thus, both signaller and perceiver were deeply invested in a cultural identity, which allowed the easy interpretation of shared knowledge. They shared a culturally learned skill, which was expressed in the output of the signallers. This is also seen with language. Variation between groups has long been argued to play a significant role in allowing the establishment of large communities based on cooperative exchange - i.e. the fact that members of a group learn the variety spoken in their group, indicates their investment in the group over others.

Roberts (2010) has presented an experimental investigation into the use of language as a marker of group identity. Participants played an anonymous economic game using an instant-messenger-style program and an artificial alien language. The game required participants to interact with other game players, exchanging resources using only an artificial language they had been exposed to. The competitiveness of the game and the number of interactions each player had with fellow players was experimentally manipulated, in order to investigate the emergence of different varieties of the alien language. It was found that, given enough interactions, participants were able to use linguistic cues to distinguish between group members and opponents. This is similar to the work of Nettle and Dunbar (1997), who showed that dialects are ideally suited to act as a mechanism to control free riders - those who take from the group without investing themselves. This is the case if and only if dialects change rapidly from one generation to the next. Roberts (2010) found that the most successful groups used what he termed a “secret handshake” - a strategy for identifying group members (Roberts, 2010). This required group members to learn and reproduce a pre-established set of actions, in order to signal group membership. This is a strategy that is very hard to fake, as non-group members were not privy to the establishment of this strategy. This suggests that the ability to learn and reproduce linguistic cues can have a major impact on group formation, identity, and membership. It also lends support to our hypothesis, that for a system of musical signals to be an honest signal and deter free-riding, it must have required a significant investment in cultural learning.

One potential criticism which could be raised is that of the importance of group membership - why would hominins need to be part of a group in order to adequately use the LSP? This can be answered by referring to the importance of the group for *learning*. This is discussed in detail in section 2.2 on page 45. Within a society in which the LSP works as a reliable indicator of cognitive abilities, both producer and perceiver must be party to the same context, part of the same cultural group, in order to learn to use and interpret culturally appropriate signals in the LSP. In order for a perceiver to tell the difference between a signaller who has invested in learning, and a signaller who, through some chance mutation, may have an innate ability to create a particular musical output, the key distinguishing factor must be that the system used changes in response to cultural changes. The fact that the LSP is culturally transmitted and can change over time means that it is stable and robust against an individual who may have this ability innately coded. Were genetic encoding the case, the genes would have to track and mimic the changing culture, potentially within the lifetime of an individual. So, a culturally-transmitted signal forces individuals to use social learning to employ it as a signal, which makes the signalling system honest. As the LSP must be culturally learned, a social group must be in place for this transmission to work effectively. As Ritchie and Kirby (2006, 290) found using computational simulations of birdsong, "where…a learning mechanism is available to the population, the genes need not code for a song explicitly as an agent can rely on obtaining a copy of the 'correct' song via cultural transmission".

Evidence from birdsong again supports this view - after ten days of tutoring with songs of their own or another subspecies, birds give stronger call responses to songs of their own subspecies, irrespective of whether they were tutored with these songs. They are also more discriminating in distinguishing between different dialects of their own subspecies (Nelson, 2000). This preference for birds' own subspecies' song may be reflected in the LSP, as group membership facilitates the reliable interpretation of the abilities of each potential collaborator, due to the signal reflecting culture-appropriate norms. One intriguing alternative is the idea that the pressure to learn the system led, as a *side effect*, to group coalitions. However, as we have previously discussed the importance of group interaction to the proposed users of the LSP (it was the newly co-dependent social groups which led to a pressure to signal abilities), we might temper this slightly by stating that the pressure to learn culturally may have led to *stronger* group coalitions.

Theories of cooperation between non-related individuals often hinge on the expectation of an ongoing relationship to sustain reciprocity (Trivers, 1971; Axelrod and Hamilton, 1981). Direct reciprocity ("You help me, I'll help you") may be key to cooperation between small groups. However, indirect reciprocity, ("I'll help you, and someone else will help me/ you will help someone else"), a more complex form of reciprocity, may be central to the development of larger groups. Nowak and Sigmund (2005, 1291) state that the "evolution of cooperation by indirect reciprocity leads to reputation building. . . and complex social interactions with ever-increasing cognitive demands". Thus, group membership may have supported, and in turn been strengthened by,reciprocity, allowing cooperation and the reliable evaluation of signallers' quality. As (Nowak, 2006, 1291) states, indirect reciprocity requires information storage, and monitoring of all individuals within the social network. These abilities may have emerged with the increasing size and complexity of social groups, which led to the pressure to signal abilities. Thus, reciprocity and group membership provided a distinct benefit to each group member. This may address the potential criticism that time spent within a cultural group does not necessarily lead to a will to cooperate - indirect reciprocity leads to an *expectation* of cooperation between group members. This would support the use of the LSP, as individuals could assume that each potential collaborator will successfully cooperate with them, allowing them simply to interpret the musical output from the LSP, and thus establish the quality of the signaller as a potential collaborator. The very fact that an individual were a group member could act like the "secret handshake" discussed by Roberts (2010) - group membership itself indicated that the individual would cooperate. Thus, group members need take no risks on non-cooperating individuals. This would serve to keep interactions within the group, as group members could avoid a potentially costly failure to successfully collaborate. So, we may state that group membership signals *likelihood* to collaborate, and musical output signals *quality* as a potential collaborator. Group membership facilitates song learning, through cultural transmission. This further supports the importance of the group to the development and use of the LSP.

As proposed in this thesis, the signallers using the LSP required both cognitive abilities relevant to musical learning, and an investment in the cultural group. The discussion above addresses the cultural investment as a potential example of a handicap, which may have served to keep the LSP system honest. However, both the investment in culture *and* cognitive abilities must be accounted for. Both elements are important to the perceiver, who needs a potential collaborator who has cognitive ability and has made a cultural investment. So we might examine the relationship between the signaller's musical output and their cognitive ability as an example of an *index*. This is another mechanism which ensured the honesty and reliability of the LSP. This follows Maynard Smith and Harper (1995, 308), who define an index as "a signal that is physically associated with a quality of interest to the receiver". The authors discuss an index as an example of an "evaluation signal" - this signal alters the information available to the perceiver, allowing the perceiver to reliably evaluate the signaller. An index is also an example of a "self-reporting signal", which provides information about some property of the signaller (Hasson, 1994). Maynard Smith and Parker (1976) used the term "assessment signal" in place of "index", and this term highlights nicely the role of the index in evaluation of the signaller.

A clear example of an index is tiger scratches on tree trunks. The height of the scratches is physically associated with the size of the tiger, and thus cannot be faked. An index can be differentiated from a cost-added signal as, while both feature correlations between the signal and a relevant quality of the signaller, only in an index is the correlation a result of physical necessity. As Maynard Smith and Harper (1995, 309) state, "in the case of an index... it is not possible for the signaller to cheat". The definitions given here seem to conform to the LSP. As proposed earlier, music-relevant cognitive abilities are said to underlie musical learning abilities. Signallers could not learn musical sequences if they did not have the requisite music-relevant cognitive abilities. This is a direct physical relationship between these abilities and musical learning abilities. So, the musical output of signallers may act as an index of their music-relevant cognitive abilities. This was unfakeable, in the same way that a tiger cannot scratch higher up a tree than it can reach. As the cognitive abilities of each individual constrained the musical signals that individual could produce, these were indexical certainly of music-relevant abilities, which may be in turn indexical of general cognitive abilities. Ritchie (2009, 80) writes that an indicator trait may be linked to the phenotypic quality of a male, in which case the female will get direct benefits. It is important here to note that Ritchie (2009) is discussing birdsong, and thus sexual selection, whilst this thesis deals with social selection. The direct benefits mentioned by Ritchie (2009) can be seen in the LSP, where musical outputs provide an index of the cognitive abilities of a signaller. The perceiver can use this to inform collaborative partner choice, and if successful, both signaller and perceiver get direct benefits, the results of the collaborative task (for example, hunting).

Thus, we can see that honesty of the signal may have been maintained by both costs (investment in cultural group) and the indexical relationship between music-relevant cognitive ability and musical learning ability. Both mechanisms worked in tandem, with the indexical relationship between signal and abilities key to the grounding of the LSP in honesty. Thus, the LSP was honest and reliable as a signal of the cognitive ability of an individual signaller. But how might the LSP have evolved? Selection is a key question when examining evolutionary theories. The following section will examine potential selection mechanisms for the LSP.

## 2.5 SELECTION

"Through selection, a set of entities, a population, will gradually adapt in response to a complex of environmental factors". So summarised Hodgson and Knudsen (2006, 478). Natural selection is so powerful that "through a blind and goalless process of differential fitness of alleles, it can gradually create structures that appear well designed for their environment" (Nettle, 2009, 84). As categorized by Bipsham (2009), theories of music and/or language evolution fall under sexual selection or mate attraction (Darwin, 1871a; Okanoya, 2002; Mithen, 2005; Miller, 2000; Merker, 2000), natural selection (Hagen and Bryant, 2003; Hagen and Hammerstein, 2009; Dissanayake, 2000; Falk, 2004), and group selection (Brown, 2000; Cross, 2009) headings. However, by far the most widely discussed mechanism for the evolution of language from a musical precursor is that of sexual selection or mate choice. Stemming from Darwin, and supported by Jespersen, Miller, Okanoya, Mithen, Merker and others, this mechanism is proposed to account for the shared features of music and language. This thesis takes a broader view, examining social selection as the potential mechanism by which the LSP evolved. This hypothesis predicts that social, rather than sexual selection, may be the key selection pressure acting on this musical signalling system. It is predicted that this signalling system was used as a cue to cognitive ability in a collaborative endeavour - hunting, for example. Thus, a pressure for the ability to choose social partners must have developed, and acted upon this system. The following sections will address both sexual and social selection mechanisms, in order to defend the evolution of the LSP via social selection. Chapter five includes a study which was designed to test whether mate or collaborative partner choice was the key pressure operating on this system. This study also serves as a test of whether perceivers of a musical signalling system really can pick up on the cues theoretically signalled by musical sequences.

### 2.5.1 Sexual selection

This thesis follows the standard definition of sexual selection as a process through which reproductive success results from differential success in competition for mates. Darwin distinguished between selection for survival and selection via competition within a species for mates (sexual selection) (Darwin, 1871a). He proposed that while survival is the key first step in existence, reproduction is an equally important second step. So, for example, some birds' brightly coloured plumage makes them more easily spotted by predators, but has the benefit of attracting mates (Fitch, 2010, 40). An asymmetry in displays was explained by an asymmetry in choice - females were the choosers. Since Darwin's discussion of sexual selection, this selection mechanism has become dominant in theories of musical protolanguage - being heavily informed by behaviour we see in birds and other so-called "musical" animals.

Geoffrey Miller is a strong advocate of sexual selection as the mechanism for the evolution of music and language, providing numerous unsubstantiated arguments to support his claim. This was discussed in detail in chapter one. Bjorn Merker also hypothesizes that sexual selection may have been the pressure which led to the emergence of language, stating; "It only remains to suggest that those early hominins who eventually gave rise to Homo engaged in such synchronous vocal signalling for mate attraction" (Merker, 2000, 318).

Sexual selection is here rejected as the sole selection mechanism acting on the LSP, as sexual dimorphism is not seen in either music or language[1]. Whilst it is true that sexual dimorphism is seen less in animal species with allo-parenting (Kleiman, 1977; Ralls, 1977, 926), no sexual dimorphism is proposed in the musical precursor. Furthermore, West-Eberhard (1979, 222) states that "something is lost by stretching the concept of sexual selection in order to make it suit new purposes, which, however interesting in their own right, tend to obscure what Darwin

---

[1] When I claim that sexual dimorphism is not seen in language, I mean this at its most abstract form - both males and females can and do use language. Sociolinguistic variables are not meant to be considered here. The sociolinguistic study of "male" and "female" language is an on-going research area. Unfortunately, whilst the evidence to date is not conclusive, the popular press sometimes lands upon a catchy sound-bite, usually in support of the idea that men and women use language differently. There is by no means a consensus to be drawn here. Labov's "gender paradox" identifies that whilst women adopt prestige forms of language at a higher rate than men, they also use higher frequencies of vernacular, low-prestige forms (Labov, 1990). These slight differences are likely motivated by sociolinguistic factors, and the degree to which these differences are really existent has been argued by, for example Lakoff (1973). Interested readers are encouraged to investigate for themselves, taking in arguments from Labov (1990); Lakoff (1973); Tannen (1990) and others.

was trying to say".

A further null finding in the area of sexual selection in relation to music comes, unusually, from a recent furore over a paper which claimed to provide evidence that dancing reveals symmetry in males (Brown et al., 2005). Stating that "dance is believed to be important in the courtship of a variety of species, including humans" (Brown et al., 2005, 1148), the authors aimed to study the relationship between symmetry and rated dancing ability. If we accept the claim that dancing is important in courtship, then we might conclude, based on this paper, that symmetry and (at least some form of) musical ability are related, and undergo sexual selection, with the dancing acting as a signal of quality to perceivers. Indeed the authors claim that "dance in Jamaica seems to show evidence of sexual selection and to reveal important information about the dancer" (Brown et al., 2005, 1148). However, this paper has been torn asunder by Trivers et al. (2009), led by one of the co-authors of the Brown et al. (2005) paper. A statistical reanalysis revealed that data had been systematically manipulated in order to present the results required. Participants were pre-screened prior to participation, and symmetrical individuals who danced badly, and asymmetrical participants who danced well, excluded. Additionally, rated dancing ability scores were averaged to inflate findings, making some previously non-significant findings significant. And so, the Brown et al. (2005) paper failed to present support in favour of sexual selection. Further critique of the sexual selection theory will be provided in chapter four, with the results of an experiment examining the relationships between cognitive and musical abilities, and physical qualities.

### 2.5.2 Social Selection

Social selection offers a much-overlooked alternative to sexual selection. The use of the term "social selection" follows the usage of West-Eberhard (1979) and Nesse (2007, 2009). As West-Eberhard wrote;

> "Social selection is differential reproductive success (ultimately, differential gene replication) due to differential success in social competition, whatever the resource at stake" (West-Eberhard, 1983, 158).

In other words, social selection is a subtype of natural selection, in which social choices made by individuals influence the fitness of themselves, and of other individuals. As defined by Tanaka

(1996, 512), it is “a selective force that occurs when individuals change their own social behaviours, responding to signals sent by conspecifics in a way to influence other individual’s fitness”. Social selection has been described as “undoubtedly one of the main evolutionary processes responsible for the emergence of both individual and group behavioural characteristics” (Crook, 1972, 264). Furthermore, as described by West-Eberhard (1979, 229-230), small phenotypic differences, under social selection, can lead to large differences in reproductive success - “one of the special characteristics of social competition is that continuous variation in a characteristic affecting social status…can have a non-linear or ‘stepped’ effect on reproductive success”. In other words, populations may split into groups of relatively reproductive, and relatively non-reproductive individuals. Fitness benefits from choosing social partners are more distant and indirect than direct reproduction, however “they can influence fitness nearly every day and at all ages” (Nesse, 2007, 149). According to Nesse, we should expect social selection to shape traits that make an individual preferred as a social partner, including;

> “(a) High levels of resources (health, vigor, personal skills, powerful allies, status, territory, and other resources), (b) Tendencies to share those resources reliably and selectively with relationship partners, (c) Accurate intuitions about what others are seeking in a partner, and (d) Strong motivations to please partners and other in-group members” (Nesse, 2009, 143).

As mentioned above, selection of the LSP is proposed to have been driven by individual level collaborative partner choice. Individuals could use the cues to cognitive ability held within musical output to make informed collaborative partner choices. Producers who were skilled at learning a musical sequential precursor were thereby signalling their fitness as a collaborative partner - they were cognitively able, flexible, well able to buffer stresses. So, those who could signal well were selected as collaborative partners, leading to increased resources (the spoils of the collaborative endeavour). These increased resources allowed these individuals to survive longer and have higher chance of reproduction, and also allowed them to lavish more resources on their offspring. Due to this increased fecundity and healthier offspring, we should expect to see musical learning ability and related music-relevant cognitive ability to appear in subsequent generations. So, we have an example of differential reproductive success as a result of differential success at a social task. The LSP is proposed to have undergone social selection, as a result of individual level collaborative partner choice, resulting in differential success in social tasks.

Due to this increased fecundity and healthier offspring, the increased music-relevant cognitive abilities, and related general cognitive abilities, were reproduced in later generations. Thus, social selection led to the increase in subsequent populations of the genes underpinning musical learning ability (due to its role as indicator of cognitive ability). This led to increased use of the musical signals as an honest indicator of cognitive ability. As Nesse (2009, 145) puts it, "preferred partners got advantages, those who preferred the best partners got advantages, and the preference and displays ran away to shape extreme prosociality that is otherwise difficult to explain". Crucially, this is a prosocial, interactional theory - as Wolf et al. (1999, 255-256) wrote, "factors other than one's own phenotype may affect an individual's fitness...individual variation in fitness can be attributed to variation in value of traits expressed by an individual's social partners".

When choosing a collaborative partner, individuals may choose different abilities to those chosen in a sexual partner. When I choose a collaborative partner, I care about his or her *cognitive* ability - ability to plan, think ahead, be consistent and predictable in decision making. It does not matter to me that this cognitive ability is supported by developmental stability or genetic quality/robustness, as our collaborative task will not result in mating. If, however, I am choosing a sexual partner, developmental stability is relevant, as my offspring will inherit the genes of this partner. For a purely *social* endeavour, genetic robustness is irrelevant, and so we might predict differential responses to signals when individuals are choosing partners for different tasks. This will be further addressed in study three.

Social selection as defined above is a sadly underdeveloped field, despite the fact that collaborative interactions are unusual to humans. That is not to say that collaboration is by any means unique to humans (Jaeggi et al., 2010) but it is found abundantly in our species. This makes collaboration an interesting area from a human perspective. As Kirby (2012, 98) writes, "in forming coalitions for joint activity, accurate assessment of others' abilities may be invaluable".

Social selection need not be seen as contra Darwin's theory of sexual selection, rather complementary. As West-Eberhard (1979, 222) writes, "when Darwin wrote about sexual selection, he focused primarily on social competition for mates...and he explicitly excluded characters involved in non-social competition for mates". As this thesis does not consider competition for *mates*, it need not be seen as 'anti-sexual selection", rather a suggestion that social and sexual selection might be in operation on the same cue. Indeed, as Nesse (2009, 140) writes, "recent use of social selection as an alternative to sexual selection is confused…social selection is not an alternative to sexual selection but the superordinate category of which sexual selection is a subtype". This thesis will focus on nonsexual social selection.

## 2.6 DOMESTICATION

As discussed earlier, this learned sequential precursor is proposed to have consisted of a system of complex sequences of musical sounds. But how might this complexity have come about? The answer may lie in domestication. Okanoya (2002) has noted that the domestic Bengalese finch has a more complex song than its wild progenitor, the White-Rumped Munia. Further evidence shows that, whilst the Bengalese Finch can learn the song of the Munia if cross-fostered, the same does not hold for the Munia fostered by Bengalese Finch parents, indicating a different role of learning in youth. In the relatively short time since the Bengalese Finch was bred from the Munia, these changes in behaviour are radical (Ritchie and Kirby, 2005), and even more fascinating when we consider that the domesticated Bengalese Finch was not bred for its song, but for its plumage. Domestication would appear to have a profound effect not only on the complexity of the bird's song, but on the neural mechanisms underlying learning. Okanoya (2002) argues that the increase in complexity was driven by sexual selection, and that domestication freed the Bengalese Finch from pressures of living in the wild. Thus, the only selection mechanism remaining was sexual selection, which led to an exponential increase in complexity.

Recent work to come out of Okanoya's lab (Kagawa et al., 2010) identified three groups of White-rumped Munia, with different levels of sympatry (shared environment with other bird species). Rate of sympatry was lowest in Huben and highest in Taipei, with intermediate levels in Mataian, in Taiwan. The complexity of the male song was analysed by song linearity (song linearity and complexity have an inverse relationship). Song linearity is here defined as the number of unique song notes divided by the number of unique note-to-note transitions. It was found that the complexity of song decreased with increased sympatry - in other words, the more diverse the flock, the more simple the song, and vice versa. The authors interpret this to mean that species identification is a pressure that maintains song simplicity. When this pressure is relieved or removed, complexity can ratchet up. They call on Deacon's theory of masking, whereby an environmental change masks a selection pressure (Deacon, 2003).

Deacon (2003) discussed the concept of masking in relation to humans. He uses the term masking to refer to "shielding or protecting something from natural selection" (Deacon, 2003, 92). He claims that both masking and unmasking have been involved in language evolution, and suggests that as this process of masking and unmasking operated, "novel selection pressures unmasked selection on previously 'neutral' variants and created advantages for certain classes of mutations that might not otherwise have been favoured" (Deacon, 2003, 93). He uses the example of species-typical calls to support this claim, stating that humans have a significantly reduced repertoire of these calls compared to other great apes. He says that masking might have led to simplification and reduced specificity, recruiting their tonal, temporal and emotional components into language functions, this being seen clearly in prosodic features.

Ritchie and Kirby (2006) summarise Deacon's argument neatly - domestication shields traits from selection which allowed random genetic drift to erode innate song biases. This "allows previously minor influences, such as mnemonic biases and early auditory experiences, to have more of an effect on song structure and learning, which results in the various neural modules involved in song production and learning becoming increasingly de-differentated" (Ritchie and Kirby, 2007, 389). This action could explain the evolution of "complex functional synergies" such as the evolution of the neural mechanisms for song production in the Bengalese finch, and also, humans (Ritchie and Kirby, 2005). Crucially, Deacon's viewpoint does not rely on sexual selection as does Okanoya (2002).

Modelling work by Ritchie and Kirby (2006) aimed to model the domestication of the Bengalese Finch, thereby testing the role of sexual selection in such a theory. In this model, simulated "birds" are exposed to songs, and select a subset of these to learn (five from 50). The bird uses a learning algorithm to learn the song grammar it will use for its lifetime. Each agent has "genes" that code for an innate species recognition device (SRD), which categorizes songs as either conspecific or non-conspecific, based on the transition probabilities within. A fitness score was implemented, which calculates how often the bird recognised its own species, and how often it was recognised by a bird of its own species, using song alone as a cue. Reproduction consists of parents of a species being chosen according to their fitness scores, and their chromosomes being crossed to create a new child. Death happens after each bird in the population is sampled a fixed number of times and its songs stored for the next generation. Then this generation is removed from the population.

Following an initial period where learning, reproduction and death interact, domestication was modelled by simply removing the fitness measure. Crossover and mutation was continued. Ritchie and Kirby (2006) state that this is appropriate as, in captivity, birds no longer need to recognise their own species for mating or territory defence, as mating is controlled by humans and they live in aviaries. Results showed that after domestication, the behaviour in the population changed significantly, with repertoires growing and becoming more varied, and these songs becoming less linear. This indicates that "the songs a bird will sing have comparatively more varied note transitions" (Ritchie and Kirby, 2006, 289). Ritchie and Kirby (2006) conclude that this behaviour is a result of "the fact that the strength of the population's filters is no longer being selectively maintained" - in other words, filters such as recognition of conspecifics have been *masked* from selection. This allows genetic mutations to accumulate, allowing the domesticated species to learn from more varied sources, which relies on early auditory experience, allowing this to have a much larger effect on adult song behaviour. This finding suggests that "an increase in song complexity. . . and increased influence from early learning can arise *without* direct selection on either trait, simply through the process of domestication" (Ritchie and Kirby, 2006, 290).

Hare and Tomasello (2005, 439) discuss domestication in relation to dogs and humans, noting that dogs appear to have evolved specialised skills for reading human social behaviour. Compared to their ancestors, the wolf, dogs understand human pointing and eye gaze. Whilst the evolution of dogs from wolves has occurred at multiple time periods, we do have an experimentally indiced example of a domesticated canid - that of Siberian Foxes (Trut, 1999, 2001). It should be noted, however, that the domestication of these foxes is, necessarily, different from the proposed domestication of dogs from wolves, as Siberian foxes are not naturally domesticated. Dogs were domesticated as by-products of human life, by scavenging scraps for example, whilst these foxes have been deliberately domesticated. The "farm fox" experiment involved the breeding of the most timid foxes from a captive population. This was tested by their likelihood of approaching a human without fear or aggression. It was found, after multiple generations of breeding the most placid foxes in a population, that the foxes had evolved other traits, aside from lack of fear and aggression. They had evolved more "domesticated" traits, such as floppy ears, smaller teeth and a smaller morpho-facial structure. Similar studies have been conducted in mice and rats (Gariépy et al., 2001; Albert et al., 2008). Hare and Tomasello (2005, 441) use this to support their view that “dogs’ skills for reading human social-communicative behaviour might have initially evolved as an incidental by-product of selection for tame behaviour”. A lack of fear and aggression towards humans probably evolved as wolves began to scavenge refuse from human settlements. They propose an analogous evolution in humans -

> “One might seriously entertain the hypothesis that an important first step in the evolution of modern human societies was a kind of self-domestication (selection on systems controlling emotional reactivity) in which a human-like temperament was selected” (Hare and Tomasello, 2005, 443).

This “human-like temperament” controls things like an urge to cooperate, which is not seen to any great extent in our nearest primate relatives. This view of human self-domestication is echoed by Deacon (2003, 94), who writes “we are a self-domesticated species”, and Kirby (2012). A key proponent of the self-domestication hypothesis is Richard Wrangham, who claims that humans are a domesticated form of ape. Hare, Wobber and Wrangham have claimed that self-domestication occurred as aggression was selected against, through the penalisation or ostracisation of aggressive individuals (Hare et al., 2012). This evolution can be seen in bonobos, whose phenotypic differences from chimpanzees reflect the difference seen between wild and domesticated animals. In a similar fashion to the farmed foxes experiment, as aggression was selected against through punishment, other behavioural, physiological and anatomical traits (increase in social tolerance, non-conceptive sexual behaviour, increased pro-social behaviour, increased problem-solving abilities, more frequent reproductive cycles, decrease in canine teeth, more gracile features, reduction in cranial capacity, depigmentisation of parts of the body) began to develop (Hare et al., 2012). These traits are by-products of selection against aggression, rather than being under direct selection themselves, and together are what Hare et al. (2012) call the “domestication syndrome”. Whilst there is an obvious difference between the controlled farm fox experiments and evolution in the wild, the same developmental pathways may be activated when in the wild and captivity, meaning that the farm fox experiment really can be seen as a model for the evolution of humans through self-domestication.

But how is this relevant to the thesis in hand? In order to signal cognitive ability for collaborative interaction, a trait common to our closest relatives may have developed into a complex signalling system, the learned sequential precursor discussed above. This was due to a pressure for reliable identification of cognitive ability. The complexity in the system may have come about through self-domestication.

## 2.7 DISCUSSION

This chapter has proposed a novel precursor to language, the Learned Sequential Precursor (LSP). This system required a cultural investment (membership of a cultural group) as well as cognitive ability and investment in learning, and acted as an honest signal of cognitive ability. The balance between cultural membership and investment of cognitive capacity and time ensured that honesty was maintained. This LSP is proposed to have been "music-like", having potentially evolved from primate drumming behaviour, through an early hominin group cohesion system. With social changes, this group cohesion system evolved to the LSP. The key point to keep in mind when evaluating the proposed LSP is that we do not propose it as the sole reason for the common factors seen in both music and language. Key to this is a pressure to be able to interpret the cognitive ability of a potential collaborative partner.

One criticism that could be levelled at the LSP is the proposed necessary relationship between musical learning ability and cognitive abilities (be they music-relevant or general). This section will address evidence against the hypothesised relationship between musical learning and cognitive abilities discussed above. Modern human musical savants may be proposed as a counter-argument to the proposed relationship. These savants are highly skilled at learning music, but have not traditionally been considered to have related cognitive abilities. Savants exhibit skills in one area which are far better than could be expected, given their deficits in other areas such as general cognition. The most extreme examples are those who score very low on IQ tests, but are very talented musicians. This may be related to the prodigious memory often found in savants, which has been described as deep but exceedingly narrow (Treffert, 2009). Musical savants could thus pose a problem for the LSP hypothesis. If, in our proposed early hominin population, musical savantism were common, or became common, then the association between cognitive abilities and musical learning would break down, and the LSP would no longer be useful as an honest indicator of the cognitive abilities of potential collaborators.

In order to address this criticism, we may look at the rates of savantism in the general population. The true prevalence of savantism is unknown, however surveys of both autistic populations and those with more broad developmental disorders can provide a rough estimate of prevalence. Whilst savantism is usually thought of as being related to autism, it is estimated that 50% of cases of savantism are from the autistic population, with the other 50% coming from people with developmental disabilities. A 1978 survey (Rimland, 1978) identified 9.8% of a population of 5,400 autistic children were identified as savants, while a 1977 (Hill) study identified a prevelance of .06%. However, some issues remain with these surveys, which make cross-comparison of results difficult. The Hill (1977) study was based on care-workers' reports, and savant status was not clearly defined. Rimland (1978) used parents' responses to estimate levels of savantism in the autistic population. Furthermore, levels of autism vary according to method of diagnosis, and have shown a rising trend in recent decades, making estimating the rate of savantism in the general population difficult. Changes in diagnosis may mean that some of the developmental disabled people in the Hill (1977) study may under current practice have been diagnosed as autistic (Heaton and Wallace, 2004, 901). However, the conclusion remains that savantism is very low in the general population. Baron-Cohen et al. (2009) found autism spectrum disorders in 1% of the population of a UK-school-based population, and if we follow the Rimland (1978) finding of savantism in 9.8% of the autistic population, even at a generous estimate that leaves us with a prevalence rate of around .1%. Whilst twin studies do point towards the heritability of autism spectrum disorders (Freitag, 2007), the low levels of savantism currently estimated suggest that musical savantism should not pose a problem for the LSP hypothesis. A prevalence of 0.1% represents one savant per 1,000 people. One might view the occurrence of musical savants as noise on the signal provided by the LSP. Evidence from prevalence studies suggests this noisiness would be very low, which leaves the LSP as a honest, reliable signal. Whilst one might have reservations about the validity of looking at modern group behaviour to conclude something about early hominin groups, we should consider that modern groups are at least indicative of early human behaviour.

Further evidence against the savant "problem" for this hypothesis is provided by evidence that savant-like skills are usually accompanied by normative levels of performance on at least some subtests of standardized measures of cognitive achievement (Miller, 1999). For example, 20% of people with autism score in the average range on measures of non-verbal intelligence (Volkmar and Lord, 1999). As seen above, savant-like skills are highly represented by the autistic population, and this finding of average non-verbal intelligence in at least part of the population perhaps suggests a previous over-reliance on standardised tests that may over-emphasise the intellectual disabilities suffered by savants.

Musical savantism has provided a robust argument against the hypothesis that musical learning ability acted as an honest indicator of cognitive abilities. This thesis will continue to assume a relationship, whilst deferring the conclusion to chapter four, where experimental results from study two will shed light on this issue.

As Brown (2000) so rightly said, any theory of the evolutionary precursor of both music and language "must provide for the common structural and expressive properties that are found in music and language… it must provide an evolutionary scaffold on which music and language can evolve after a period a [sic] divergence and differentiation" (Brown, 2000, 277). Does this proposed learned sequential precursor, a precursor to both music and language, satisfy these requirements? I believe so. Complexity, as suggested by the above hypothesis, is clearly to be seen in the complex and infinite linguistic and musical combinations, whilst melody and rhythm are each seen, in varying degrees, in both music and language. Indeed, Motherese may serve as an analogy for the learned sequential precursor, as it is melodic, makes use of rhythmic regularity and variety, and dynamic variation in intensity, volume, speed, and alterations of vocal timbre (Dissanayake, 2000, 394). However, I must stress that this in an example of analogy, not homology. The following chapter will situate the musical precursor in time, and discuss possible routes this precursor may have taken in its different evolutionary paths to language and music.

# CHAPTER 3

# EVOLUTION OF LANGUAGE AND MUSIC

Let us investigate how this proposed learned sequential precursor could have evolved to language. So far we have seen that our ancestors may have used a musical learned sequential precursor to signal cognitive ability. This signalling system would have ramped up in complexity to a system of complex sequences of rhythms (vocal and manual), signalling cognitive ability, which allowed perceivers to make informed collaborative partner choices. But how did this evolve into language (and music)? The LSP is envisaged to provide a foundation of vocal sounds and movement, upon which to build speech sounds and gestures. This theory focuses on the signalling of musical sequences, which can be seen to be mostly vocally-led. However, this theory does not exclude the very rich and important impact of gesture on communication. This precursor could have been a multimodal, integrated whole made of vocalisations, gesture, dance, and movement. Importantly, once this ability to learn a sequential precursor system came on stream in humans, this facilitated a slow and gradual evolution of language, with one potential scenario outlined below.

As we have seen in section 2.1.1 (page 41), the learned sequential precursor is proposed to have required the interplay of three key elements - motor skill, cultural learning and productive creativity. Here, we will analyse the importance of these three elements in both music and language, the modern instantiations of this proposed precursor. I propose that all three elements are key to music. It is clear that great musicians have mastered all three areas - they have manual control of their instrument(s), they are strongly imbued with their musical culture, and they also exhibit great creativity - they take in many disparate pieces of information about the musical culture they are part of, and synthesise this information to allow them to create culturally-relevant, but interestingly novel, musical pieces. Key to this is a delicate balance of cultural learning and creativity - it is the very rare musician that can do something entirely creative, with no cultural grounding of his output. However, I argue that even amateur musicians are adept at using all three elements of the system. An amateur musician likely has control of his instrument, and is strongly enculturated in a certain musical tradition. Thirdly, I argue that this amateur *can* be creative with his music. Even if he confines himself to covering other musicians' creations, he is likely to add his own creative flair to the piece - he will not just copy exactly to the beat and note the previous version of the piece.

We see a similar relationship between the three elements when it comes to language. Motor skill and cultural learning are absolutely key to language acquisition and use. They might be seen as the first steps in the developmental trajectory. Everyone (aside from those with linguistic deficits) is a master of the mechanics of speech or sign, and a fully enculturated member of a culturally learned system. Additionally, creativity is seen in every novel sentence. Language is extremely productive and thus, language use is intensely creative. Whilst language does have the role of information transfer, and thus must be utilitarian to a certain degree, this does not negate the importance of creativity in language, within the bounds of the cultural, linguistic system. So, the LSP can be seen to have facilitated the evolution of both language and music. Plausible steps from the LSP to both language and music are outlined below.

## 3.1 Holistic protolanguages and segmentation

A key debate in the field of language evolution is the synthetic vs holistic argument. This debate regards whether human protolanguage consisted of a small number of referential words which were combined in short sequences (synthetic), or holistic utterances, each with defined meaning, which were, at a later stage in evolution, analysed to form words. This argument centres crucially on *protolanguage*, a later stage in the evolution of language than the LSP. Many aspects of the LSP such as complexity, sequential structure, and vocal and gestural modalities may have been facilitated by the LSP. This section is intended to be informative of the possible evolution of language from the LSP. Whilst this thesis will not address this in great detail, it is important that we are familiar with possible steps to language from the LSP.

Derek Bickerton is a key proponent of the synthetic approach. Bickerton (1981) views protolanguage as consisting of sets of words, strung together in some rudimentary grammar.

He states that this may resemble early child language, human language as used by apes, or pidgins (Bickerton, 1996). However, he does not subscribe to a continuation view of vocal communication, rather of cognitive processes, saying;

> "Since language is primarily a representational system, its antecedents are to be found not in primitive forms of communication but rather in the means by which earlier... species represented to themselves the universe they inhabited ... Protoconcepts which could serve as referents for nouns and even verbs... were in place by the time the higher primates had developed" (Bickerton, 1990, 100).

This view is supported by James Hurford, who writes "while apes may perhaps not be capable of storing such complex structures as humans, it seems certain that they have mental representations in predicate-argument form" (Hurford, 2003, 45). Thus, the assumption follows that these proto-concepts of nouns and verbs derived from primate mental representations. Word-like things based on these representations evolved before the syntactic principles to form phrases emerged.

A very early instantiation of the holistic protolanguage theory came from Jespersen (1922), in discussing his view of language's evolution from music. He draws an analogy between the evolution of writing systems from primitive art and that of language from holistic musical utterances. He writes;

> "In primitive picture-writing, each sign meant a whole sentence or even more the image of a situation or of an incident being given as a whole; this developed into an ideographic writing of each word by itself; this system was succeeded by syllabic methods, which had in their turn to give place to alphabetic writing, in which each letter stands for, or is meant to stand for, one sound. Just as here the advance is due to a further analysis of language, smaller and smaller units of speech being progressively represented by single signs, in an exactly similar way, though not quite so unmistakably, the history of language shows us a progressive tendency towards analyzing into smaller and smaller units that which in the earlier stages was taken as an inseparable whole" (Jespersen, 1922, 437).

Alison Wray (1998, 2000, 2002b,c) is a strong proponent of the holistic protolanguage theory. Wray states that "if we are seeking a language without grammar, we should not start with words, but with holistic utterances with an agreed meaning" (Wray, 1998, 49). This is because these holistic utterances do not have component words, and further, they have no regularity that can be understood by users. These holistic utterances were kept distinct by phonemic contrast, creating a "phonetically sophisticated set of fomulaic utterances, with agreed function-specific meanings" (Wray, 1998, 51). These utterances developed from, and performed the same functions as, primate calls - expressing and negotiating relationships, and inviting specific actions of others - social manipulation and mediation. Wray (1998, 50) is clear that "protolanguage developed *out* of the older holistic communication" systems of primates, and the utterances "were a direct development from the earlier noises and gestures" (Wray, 1998, 51).

These holistic utterances, in humans, may have been coined by someone, and depending on that individual's status, and the usefulness of each utterance, would or would not have been adopted by others. Their meaning would be decoded by context, pragmatics, and gestural cues. This is similar to how slang words or stock phrases spread through populations. Indeed, Wray (1998, 60) uses the existence of stock phrases ("formulae" in her terminology) to support her theory. She reports that formulae are used today to manipulate others, prove group membership and identity, and help with fluency and turn holding. She claims that these could be seen as examples of holistic phrases - the meaning of the whole is no longer composed of the meaning of its parts, but has an idiomatic or non-literal meaning. Wray (1998, 63) writes that "holistic language may be picking up on the shortfall between what we want to say and what we have the processing power to compute from scratch" by removing the burden of processing. So, we have a system in place now where both approaches - holistic and synthetic - have a place, with each taking a different role, relieving certain pressures.

Wray supports her claim for holistic protolanguage by drawing on evidence that primate calls are largely holistic in nature (Reiss, 1989). Vervet monkeys are a clear example, as primates which have three distinct calls for different predators (Seyfarth et al., 1980). Wray calls these "complete utterances with a function, namely to warn" (Wray, 1998, 50). It is true that we may never know definitively whether a vervet is conveying "snake!" or "beware of the snake in the grass" with his call. Wray uses this to claim that these calls are discrete *messages*, not discrete *words*. Wray herself explicitly supports an evolutionary relationship between primate calls and human language "the holistic cries and gestures of our pre-human ancestors were transformed, over a long period of time, into a phonetically expressed set of holistic message strings" (Wray, 2002a, 115) - a key difference between her view and that of Bickerton. Michael Arbib takes a similar, but distinct view; whilst primates have a fixed number of calls, early humans could create new utterances which were "acquired through learning within the community" (Arbib, 2003, 186) – in the same way, perhaps, that the learned sequential precursor outlined in this thesis was learned.

As briefly mentioned earlier (see section 1.1.12, page 20), Wray (2000, 294) states that segmentation of holistic protolanguage could come about as learners mistakenly infer a regularity between two holistic proto-linguistic strings (e.g. if tebima means "give that to her" and kumapi means "share this with her", then the learner might infer that "ma" refers to "her"). Thus, (mistaken) segmentation could operate on a holistic protolanguage until an arbitrary compositional language is reached. Iterated learning experiments have both modeled (Kirby, 2001), and empirically tested (Kirby et al., 2008) the evolution of language structure over cultural transmission. In these experiments, a learner is given an entirely random, holistic language. However, over cultural time, this language becomes regular and compositional. This can be seen to occur in a similar manner to Wray's segmentation theory - learners mistakenly infer structure where there is none, and so, over cultural time, this becomes amplified until the strings become segmented and compositionality occurs. If we combine the two arguments, we can see how our learned sequential precursor could become segmented and evolve into a structured language.

### 3.1.1 ITERATED LEARNING AND THE SEGMENTATION OF HOLISTIC UTTERANCES

As Mesoudi and Whiten (2008, 3490) state, "cultural transmission is a fundamental component of cultural evolution. Without transmission there can be no evolution". Thus, in order to investigate the evolution of a culturally learned system such as language, we must look to the modes of its transmission. Kirby (2001) coined the term "Iterated Learning" to describe the inter-generational transfer of cultural information such as language. So, "iterated learning is a process in which an individual acquires a behaviour by observing a similar behaviour in another individual who acquired it in the same way" (Kirby et al., 2008, 10681). Kirby was, in this paper, following a line of research that studies languages themselves as adaptive systems. In this computational model, each generation consists of an adult agent, which is given a set of random meanings to produce signals for. These meaning-signal pairs form the training data for a learning agent (Kirby, 2001, 102-103). After this, the learning agent becomes an adult one, a new learning agent is introduced, and the cultural transmission continues. Here, the signals were ordered strings of letters, and the meaning space a set of two components which varied over five values, creating a maximum of 25 possible meanings. Learning here consists of two parts - *Incorporation*, and *Generalisation*. When a new meaning-signal mapping is encountered, the agent creates a rule covering this input, and incorporates it into the grammar. Then, the algorithm tries to integrate this new rule into its existing grammar, looking for possible generalisations across data (Kirby, 2001, 103).

Results showed that a fully compositional syntactic meaning-string mapping can emerge from an initially unstructured communication system (Kirby, 2001, 102). Words or morphemes which correspond with parts of the meaning space emerge, and regular rules for the combination of these are created. This comes solely from cultural evolution - there is no biological evolution in the system. Kirby sees this an an example of language evolving to better its own survival - pressures act on the language as it undergoes cultural transmission, which result in compositional structure from what began as a holistic, un-analysable whole. These pressures can take the form of small biases - for example, each agent may have a small bias towards compositionality, which, over transmission, amplifies to a powerful effect. Current modelling work by Thompson et al. (2012) shows that iterated learning can amplify weak biases; “cultural transmission unmasks these weak biases in individuals, yielding large effects in the population’s language” (Thompson et al., 2012, 558). This model is proposed as evidence contra the nativist hypothesis, finding that “cultural transmission renders the biological evolution of strong domain-specific innate constraints unlikely, but we nonetheless witness strong universal tendencies in the cultural systems of these simulated populations” (Thompson et al., 2012, 557).

Kirby, whilst a leading force in the field, is by no means alone in using the ILM (Iterated Learning Model) to investigate language evolution and language change. Many interesting original and review papers have been published by, among others, Kirby and Hurford (2002), Smith et al. (2003), Kirby et al. (2004), Griffiths and Kalish (2007), Kirby et al. (2007), Ferdinand and Zuidema (2009)

However, one criticism that computational modelling suffers from is obvious - these simulated agents are precisely that, *simulated*. Recent work has successfully replicated the findings from Kirby (2001) in a human population of learners. In order to do so, a paradigm was adopted from psychology, that of diffusion chains (Bartlett, 1932). Under this paradigm, a task is passed from person to person to model generational cultural transmission. Like a large game of Chinese Whispers/ Telephone, the first individual learns a task. Then, their output is given to the next generation as their input, and so on throughout the transmission chain. In a landmark 2008 paper, Kirby et al. (2008) used the iterated learning model in an artificial language learning experiment. This was conducted in order to test the computational and mathematical models to date, and in doing so, expand the experimental literature in this field. Crucially, this paradigm was investigating "cumulative, adaptive, non-intentional cultural evolution" (Kirby et al., 2008, 10681) in a human population.

In the Kirby et al. (2008) paper, participants were asked to learn an "alien language", pairings of written labels with visual stimuli - made up of three dimensions, colour, motion and shape. This yields a total of 27 unique objects, and nonsense labels associated with them. In the training phase, participants were taught mappings of label and object, and later, in the test phase, they were asked to provide labels for visual stimuli presented to them. However, in order to test participants' ability to generalise, they were only taught half of the data - introducing a 50% bottleneck. Participants were explicitly told to learn and reproduce the language as best they could.

It was hypothesised that structure would emerge, over generations, due to the pressure for the language to be faithfully reproduced. So, there should be a decrease in transmission error, and the emergence of linguistic structure - regular, predictable mappings between object and label. Results of the first experiment, involving four chains of ten individuals, showed a significant decrease in transmission error between initial and final generations. This suggests that the initially random language was adapting to be more readily learned and transmitted. Further, it was found that the number of distinct meanings in the language decreased rapidly, thereby introducing ambiguity in the language. In order to counter this trend, a further experiment was conducted following the same approach. However, as a proxy for a pressure for expressivity, after each participant had reproduced the language, homonymy was filtered out. This was intended to avoid under-specification and ambiguity. Following this change to the experimental paradigm, both predictions were upheld; languages became more learnable, and became increasingly structured, with cultural transmission over time.

So, these experiments have shown that initially random languages, which were essentially holistic utterances describing a whole object on multiple dimensions, can, in just ten generations, become a regular, compositional language, in which each compositional part refers to one dimension of the meaning object. This is a result of iterated learning, with a pressure for expressivity.

A recent study by Verhoef et al. (2011, 2012) has further tested the analysis of holistic utterances into compositional parts, while abstracting away from meaning. In order to test the arguments surrounding the holistic/ synthetic debate, Verhoef et al. (2011) used the iterated learning model, subtracting the meaning element of the signal. In this experiment, Verhoef asked participants to learn and recall twelve distinct whistle sounds that had been produced using a slide whistle. The slide whistle was used in an effort to abstract the task away from a linguistic one. Participants were initially trained to produce twelve distinct whistles, before being asked to recall twelve. Following the Iterated Learning Model, participants learned the output of the previous generation. The initial whistles were holistic and did not exhibit combinatorial structure. No meaning space was used, as the effort was to model the emergence of "bare phonology".

Results showed that human learners quickly generate structure from an initially holistic, structureless set of signals. The task of learning and reproducing twelve totally distinct whistles is a difficult one, and so, learners tend to overgeneralise the structure they infer, introducing mirrored, combined, or repeated versions of whistles (Verhoef et al., 2012, 370). Thus, whistles come (over generations) to share elements, becoming increasingly compositional. These elements are akin to "building blocks", increasing the learnability of the signal set. This study further supports the findings of Kirby et al. (2008), while adding an extra dimension - the "utterances" in Verhoef et al. (2012) were continuous sounds made on a whistle, and thus, not discrete symbols that participants are familiar with. This further supports holistic protolanguage theories - "continuous, holistic signals can transform into signals built up out of discrete segments, independent of meaning" (Verhoef et al., 2012, 373). This seems to address Bickerton (2003, 87)'s suggestion that "how smaller units can be factored out from holo-phrases when even their global meanings are so potentially ambiguous remains unclear" - here, there was *no* meaning, and yet a compositional, structured system emerged.

These experimental and computational models support Wray's segmentation analysis. This thesis takes the view that the LSP acted as a precursor to a holistic protolanguage. The change from LSP to protolanguage may have come about due to domestication, as discussed in section 2.6 on page 61.

## 3.2 Music as an evolutionary fossil

Much controversy exists over whether music should be viewed as an evolutionary adaptation, merely a non-adaptive pleasure-seeking behaviour, or an evolutionary vestige. This discussion may inform, and be informed by, the LSP. The most well-known argument for the Non-Adaptive Pleasure Seeking account (NAPS) comes from Pinker (1997), who likens music to "auditory cheesecake". He uses the example of cheesecake to highlight his view of music. In pre-modern times, sugars and fats were rare in the diet, but nutritious and valuable in the small amounts available. Thus, humans became hard-wired to seek sugar and fat, and our tastes evolved to favour foods with high sugar and fat values. However, in modern times, humans have a diet which contains unnaturally high levels of sugars and fat, and it cannot be ignored that levels of heart disease, diabetes, and other health problems related to this kind of diet are rising to critical levels. While pre-modern humans' need for sugar and fat was adaptive, as it aided their fitness, this propensity to seek out fat and sugar is now harmful, as the environment has changed. So, our greed for sugar and fat is thoroughly non-adaptive. Pinker states that music is like auditory cheesecake - a food item with large amounts of sugar and fat - as it serves us no adaptive advantage.

However, Huron (2001) argues against the analysis of music as a NAPS behaviour. He states that if music really had no survival value (i.e. were non-adaptive), then any predisposition towards this behaviour would actually *worsen* one's survival - spending inordinate amounts of time, money and energy in enjoying music would lessen one's likelihood of survival (Huron, 2001, 46). This does not seem to be the case, and indeed, some hunter-gatherer groups have members whose sole purpose is that of music maker. These people are supported by the rest of the community, as their role is seen as so vital to the group. Following from this argument, Huron (2001) states that music should be a recent phenomenon, as otherwise music lovers should have become extinct long ago (due to the non-adaptive effects). However, examples such as the Divje Babe flute, dated to roughly 43,100 years ago (Fink, 2002) and the many other historical examples of music instruments, counter this argument.

Another argument for the existence of music in its modern form is that it is evolutionarily vestigial. In other words, although music may at one time have conferred an evolutionary advantage, it now remains as a vestige of its former function. As outlined in this thesis, if our premodern ancestors used a musical sequential signalling system, those who were skilled in the use of this system were likely to have increased fitness. Thus, sequence learning skill conferred an evolutionary advantage. However, over time this signalling system evolved in one way into language, and in another into music. Thus, music was no longer used or needed as a signalling or communicative device, and became vestigial. As Huron (2001, 43) suggests, in addressing possible evolutionary origins for music we must ask "what advantage is conferred on those individuals who exhibit music behaviors over those who do not?". In looking at music's role today, we should not just look at direct fitness. Rather, we should look at the many advantages musical behaviour provides us with.

Charles Darwin described music as a remnant of a pre-adaptation, the proposed musical protolanguage detailed in this thesis. This view has been adopted by, amongst others, Brown (2000) and Mithen (2005). If we adopt this view, we cannot view music as an adaptation, as its original function has been taken on by language, nor an exaptation, as it was once honed to a particular function. Neither can music be viewed as a vestigial trait, as it still imposes physiological costs on the producer, and is implicated in emotional affairs and mating and socialization, thus retaining the power to have huge effects upon our physiology.

If music were truly useless, we should expect to see it selected against, like the eyes of a cave-dwelling fish. However, music is not selected against, and plays a huge role in modern human life. It must be viewed as "an ex-adaptation, a fossil of a former evolutionary regime, a biologically grounded cognitive domain whose current use is neither the same as its original function, nor completely different" (Fitch, 2006b, 86). This "former evolutionary regime" may well be the LSP, a "biologically grounded cognitive domain" - music is not the same as, nor completely different from, the LSP, but shares some key similarities, presumably from evolutionary progression.

## 3.3 EVOLUTIONARY TRAJECTORY OF THE LSP

This thesis proposes an early precursor to protolanguage, a learned sequential precursor (LSP) made up of complex musical sequences. This system was used to signal cognitive ability to perceivers, who need not themselves be producers of the system. This musical precursor would have acted as an honest signal of ability, kept such by the investment required. We can envisage an evolutionary trajectory in which this learned sequential precursor evolved to language. As discussed by Hagen and Bryant (2003), music is a powerful coalition tool. As this proposed precursor to language was used as a cue to fitness for collaborative tasks, we may see how it might have evolved out of a coalition/coordination/bonding system (the "Group Cohesion" element; see Figure 1.1 on page 3). This coordination system may have been in place due to increased group living among our ancestors, and a seemingly human-unique drive for group cooperation. The learned sequential system may have evolved biologically through selection for honest signalling, perhaps due to changing environmental and social pressures associated with Homo erectus.

Once the learned sequential precursor was in place, it would have remained stable as an honest signal of cognitive ability for collaborative endeavour, until some external pressure or environmental change altered its path. In order for this subconcious signalling system to change to a conciously controlled, communicative system (a change from the LSP to a protolanguage), there must be both an ability to infer communicative intent in others, and the motivation to engage in this behaviour (Kirby, 2012, 101). If these requirements were in place, this complex learned signalling system could have been appropriated for the communication of meaning. Signalling of abilities would have been lost from the system at the same time as communicative pressure bore down on it. A pressure for expressivity would have been inherent in its new role as communicative system (as seen with the introduction of semantics in Figure 1.1).

This communicative system is envisaged as a collection of holistic utterances, with no compositional meaning. This is when cultural systems, building on the biological changes such as increased cognition, would have facilitated the evolution of language. As discussed above (section 3.1.1, page 72) both computer models and human experiments have highlighted how cultural processes (cultural transmission) can act upon holistic strings, breaking them down to ordered, compositional, regular word-like parts of language. Thus, holistic utterances could have become compositional words. The development of this kind of structure seems to be driven by learnability - a holistic system, in which each utterance must be learned independently, will quickly become saturated and exceed human's memory reserves. A smaller system with a finite number of repeated elements, such as we see in language, is predictable, learnable and productive.

Once a regular, compositional system was in place, or indeed at an earlier stage in the system, the musical element may have been downgraded, to the kind of "musicality" we see in prosody now. Cultural transmission mediated the "arms race" between learnability and expressivity, creating a syntactic structure that was reflective of the semantic structures being conveyed (Kirby, 2012, 102). Thus, language could have evolved from a sequential signalling system, through a protolanguage, which may be envisaged as described by Wray (1998, 50), a "successful, if limited, system of vocalisations and gestures that. . . oiled the social wheels and achieved a number of specific interactional goals".

The musical elements of the LSP, divorced of any pressure to signal fitness, would have developed along a different trajectory to that of language, developing its own "floating intentionality" (Cross, 2003), and retaining emotional aspects of an earlier group cohesion system. Divorced of expressive semantic pressures, it could have "run away" with emotion, laying the foundation for the emotion found in music. This development from LSP to music may have passed through a protomusical stage, before settling as music - a complex, learned, culturally transmitted system.

The following chapters will outline experimental work into this proposed musical precursor to language. Chapter four presents two experiments, the first examining the role of complexity in the proposed musical precursor. The second is a large-scale experiment correlating musical learning ability and cognitive ability. Chapter five compares two possible drivers of this system - mate choice and collaborative partner choice, in order to address social and sexual selection. This chapter continues with a discussion of when we might expect this musical sequential signalling system to have first emerged.

### 3.3.1 Situating the proposed musical precursor in time

This musical signalling system is proposed as a very early stage in the evolution of language. This section will attempt to situate the emergence of this system in evolutionary time, drawing on the archaeological record and what we know about the evolution of the genus Homo. In order for a musical signalling system to be in place, the users of such a system must have had an appropriate social structure, that of group living and interdependence between individuals - this a result of pressures from harsh environments, limited resources and large amounts of predators. They must have been able to form basic concepts, to plan (to the extent that one needs to hunt) to decode cues in a musical signal, and, crucially, to learn a complex system, one based upon vocal sounds alongside bodily movements. So, vocal control must have been in place, albeit not necessarily resembling the sounds modern humans make. We can attempt to situate the musical precursor by drawing on what we know of human evolution, from the archaeological record. The aim of this section is to locate in time the *earliest possible* emergence of a musical signalling system, based on the prerequisite abilities. This may inform future discussion of the evolution of language from a plausible musical precursor.

#### 3.3.1.1 Evolution and attributes of the genus *Homo*

"I think that there is probably no other issue more important to the general public than the evidence for the first appearance. . . of our own genus. We are special, we are different, and we are concerned" (Leakey, 2009, 4). So wrote Richard E. Leakey, son of Mary and Louis Leakey, famed archaeologists responsible for the discovery of the tools and fossils of early hominins at Olduvai Gorge, and most notably the Laetoli footprints, hominin footprints which were preserved in volcanic ash. Richard, himself a respected archaeologist, went on to write "I believe that species designation as it pertains to our understanding of the evolution of the genus *Homo* is obviously an important and sensitive issue that we need to think about" (Leakey, 2009, 5).

If we look at the genus Homo in more detail, we find that "it is widely assumed, but rarely articulated, that the species included in Homo form a monophyletic group, and it is also widely assumed, but also rarely articulated, that the taxa all share functional characteristics or competencies" (Wood, 2009, 18). It is widely agreed upon that these functional competencies include the ability to make stone tools, and the ability to hunt.

Darwin, in 1871, proposed an evolutionary scenario linking bipedal locomotion, canine reduction, tool manufacture and brain enlargement at the base of the human lineage. This was the defining scenario for decades, until by the 1940s, research into Australopithecus proved that bipedality and reduced canines emerged well before encephalisation and any tool record. So, from this point onwards, large brains and tool use and manufacture "were elevated to a foundational role in the divergence of the Homo lineage from some generalised australopith form" (Kimbel, 2009, 31). What we do know is that the earliest well-sampled, well-attested record of fossil hominins bearing characteristics found *only* in humans date from c. 1.7 - 1.8 million years ago (Kimbel, 2009, 31), and furthermore, by 1.5 million years ago, there is circumstantial evidence linking Oldowan tools, stone-tool cut mammal bones, hominin fossil remains and evidence for hominin encephalisation beyond that of the Australopiths (Kimbel, 2009, 31). Kimble concludes that

> "If we grant that the co-occurence of encephalization, stone tools and paleontoligical evidence for meat consumption is functionally meaningful, is it reasonable to take these presumptively derived differences between Homo and the australopiths as explanatory of the origin of the lineage leading to modern humans?" Kimbel (2009, 31).

The early stages in Homo evolution were peopled by Homo habilis, Homo rudolfensis and Homo erectus. Homo habilis retains some primitive features which link it to Australopithecus, and little is known about Homo rudolfensis other than that it had a relatively large brain and that it overlapped geographically and chronologically with other Homo species (between 1.9 and 1.8 million years ago). Homo habilis evolved from Australopithecines, with some debate over which species of Australopith exactly was ancestral (Dunsworth, 2010). It is widely accepted that Homo habilis is the common ancestor of all of the later Homo species, including Homo sapiens (Dunsworth, 2010, 353), and it has been dated from 2.4 - 1.4 million years ago. H. rudolfensis dates from 1.9 million years ago. Homo erectus, dating from 1.8 (when it diverged from H. habilis) to 0.4 million years ago (Dunsworth, 2010), may represent the beginnings of what most lay people would accept as "modern human" - it was the first obligate, full biped, and it was also the first to disperse outside of Africa, with remains recovered from Eurasia and Indonesia (Dunsworth, 2010, 354). Homo erectus "signifies a major shift in hominin evolution, most notably through increased brain and body size and increasingly complex tools and behaviours" (Dunsworth, 2010, 353).

Homo erectus was the first of the modern Homo species. It thrived for over a million years, from 1.8 to 0.4 m.y.a. It was the first to leave Africa, and had a larger cranial capacity than the two other early Homo species. However, it is unclear what exactly precipitated its extinction, other than the emergence of Homo Neanderthalensis and Homo sapiens, with their larger cranial capacity and more sophisticated tool technology (Dunsworth, 2010, 355). We are fortunate to have a rich fossil record for Homo erectus, which allows us to draw inferences about its behaviour. There is some debate over the terminology used to describe this species of Homo, with some scholars using ergaster in place of erectus. Homo ergaster refers to the earliest members of Homo erectus which tend to be smaller and are found largely in Africa, so here, I am following Dunsworth (2010) in using erectus to refer to all members of the species, be they early or late.

With (later) Homo erectus real encephalisation began, with hypothesised pressures of environment, predator evasion, social living, opportunities created by dextrous hands which were freed by bipedalism, increased meat and marrow in the diet (Dunsworth, 2010, 359). This increased brain size is associated with increased behavioural and cognitive complexity. In particular, Homo erectus is associated with Acheulean tools, more complex than Oldowan, and exhibiting tear-drop shaped bifaces known as handaxes. These were probably used to process plant material and skin and to butcher animals (Dunsworth, 2010, 359). The transition between Oldowan and Acheulean tool manufacture is dated to approximately 1.7 million years go. This phase of "techno-cultural evolution" is associated with the emergence of Homo erectus (Roche et al., 2009, 137). Also associated with this time period is a broadening of geographic range and an increase in number of identifiable Acheulean sites throughout Africa (Roche et al., 2009, 137).

As well as, and no doubt related to, more complex stone tool manufacture, Homo erectus also shifted to eating more meat. This deduction is based on anatomical evidence that molars in this species are steeper than in Australopiths, which suggests teeth more useful for shearing meat or breaking up tough foods (Ungar, 2004; Ungar and Scott, 2009). However, dental microwear evidence does not suggest that early Homo was eating tough foods (Dunsworth, 2010, 360), and so we may conclude that Homo erectus was eating more meat than previous Homo species, albeit in a variable diet according to the variable environment. "Brain enlargement, reduction in molar tooth size, increased stature and other features of early Homo did not evolve in a vaccuum" (Roche et al., 2009, 135). These evolutionary changes are as a result of complex interactions between early Homo and other hominin species, and between Homo and the environment. Environmental changes in particular are determined to have been crucial in the evolution of Homo as a genus. These environmental changes were a result of global climate changes, creating a cooler Plio-Pleistocene climate which opened up the savanna grasslands. These grasslands were populated by large herbivores, a source of meat for Homo erectus (Dunsworth, 2010, 360). There is evidence to link the spread of Homo erectus to more increased grasslands, and more open and wetter landscapes (Bobe and Leakey, 2009; Sponheimer and Lee-Thorp, 2009). This is presumably linked to increased resources. Roche et al. (2009) has discussed the fact that stone-cut-marked bones increased in number around the time of Homo erectus, in association with Archeulean artifacts. Butchery marks suggest that Homo erectus was extracting flesh using sharp-edged flakes and marrow using rounded hammerstones (Roche et al., 2009, 140). Cut marks from stone tools are clearly distinct from teeth marks from large predators, so we can conclude that Homo erectus was butchering and eating larger amounts of animal meat than ever before in the Homo lineage. As (Roche et al., 2009, 140) puts it, "traces of hominin feeding provide paleoanthropology with a more specific itemization of dietary elements than is currently possible from other lines of evidence". This is in contrast to Olduvai sites, which have scarce evidence of butchery-marked bones. This "might be a signal that large mammal carnivory was infrequent during the early stages of this adaptation" (Roche et al., 2009, 140). So, it may be the case that only with the advent of Acheulean tools did Homo begin to habitually eat meat. Roche proposes that

> "It is during the Oldowan that members of presumably one hominin lineage expanded their range of interactions with large carnivores from sole status as prey, to include indirect and/or direct competition with at least some species for larger mammal carcasses and/or live prey, eventually leading to modern humans' status as the top predators in most ecosystems" Roche et al. (2009, 140-141).

We can reasonably suggest that the more complex stages of this evolution - competition with species for mammal carcasses or live prey - emerged with the more sophisticated tools and increased brain power of Homo erectus.

Linked to this increased time spent in the savannah, coupled with endurance running in intense heat (Lieberman, 2009), is a pressure for temperature regulation, and so it is hypothesised that this is the point at which Homo lost body fur and evolved intense sweat cooling (Jablosnksi, 2004).

Dunbar (2003), in his *Social brain* hypothesis, addresses both the change in diet and social environment in primates. This hypothesis was proposed to explain primates' unusually large brains, in particular the neocortex. It suggests that "the cognitive demands of living in complexly bonded social groups selected for increases in executive brain (principally neocortex" (Dunbar, 2003, 163). He mentions Byrne's (1988) suggestion that primates and non-primates differ crucially with regard to social skills, as a key point in the development of alternative hypotheses for primate brain evolution - this was the point at which researchers took an interest in social skills. Dunbar, in his work on social group and neocortex size in humans and primates (Dunbar, 1992) found a neat correlation between neocortex size and group size in non-human primates, and used this to extrapolate the ideal social group size for humans, based on our neocortex size. He predicted a group size of 148, which he claims represents an intermediate group size between the overnight camps of foraging groups (typically 30 - 50 people) and the size of tribal units (1500 - 2000 people). Furthermore, he states this is roughly equivalent to the set of individuals with whom one can have a personal relationship (Dunbar, 1992; Hill and Dunbar, 2003) - the number of people one could ask a favour of, and expect to have it granted. Dunbar claims that this is supported by analysis of average Hutterite community size, army company size (typically 180), and also by research specialisations in the sciences, which he claims typically consist of up to 200 individuals but rarely more (Aiello and Dunbar, 1993). All of these numbers fall within the 95% confidence limits for group size as predicted by relative neocortex size.

In other words, the larger the neocortex, the larger the group size. A larger neocortex allows the individual to deal with a larger number of conspecifics, understanding the interactions between them. Dunbar (2003, 169) states that "when the ability to maintain large group sizes is at a selective premium, it may be necessary to enlarge the whole brain in order to be able to produce the enlarged neocortex required to support large social groups". This explains the unusually large brain found in primates, and the further enlarged neocortex. However, as Dunbar himself addresses, this argument that a large neocortex supports large social groups does not explain a pressure for the *growth* of larger brains and neocortex. Large groups incur significant costs, in terms of ecological, reproductive, and energetic costs (Dunbar, 2003, 169), and so, large groups cannot be simply a by-product of having a large brain, as "the costs of living in large groups would inevitably result in their rapid dispersal if there were no intrinsic advantages to living in large groups to offset these costs" (Dunbar, 2003, 169). Dunbar uses this line of reasoning to conclude that "parsimony and biological common sense would suggest that it is group size that drives brain size evolution rather than brain size driving group size, and that group size is itself a response to an ecological problem (most probably predation risk)" (Dunbar, 2003, 169). This created a "social brain" upon which the change to a meat-based diet could act, providing massive encephalisation in Homo erectus. This again fed into hunting abilities, therefore creating a ratcheting effect of brain size and social development, and providing an environment in which a musical learned sequential precursor to provide cues about collaborative fitness would flourish.

Aiello and Dunbar (1993), used the same approach to extrapolating group size from neocortex size as used in Dunbar (1992), this time extrapolating mean group size for hominin species. Although no neocortex volume measurements exist, the authors argue that neocortex size can be estimated from total cranial volume, using a scaling relationship between brain parts identified by Finlay and Darlington (1995). The resulting extrapolated group sizes are bracketed by those of chimpanzees and humans. The data from Homo erectus predicts a group size of between 90 and 120, leading Dunbar (2003, 173) to conclude that "group size probably remained within the broad range for living great apes until well into the Homo erectus period and only began to rise significantly above this level from about 1 MY".

Dunbar proposes three possible reasons for increased group size in hominins. The first suggests that the habitat the hominins occupied presented "proportionately higher risks of predation" than those occupied by babboons, macaques or chimpanzees (Aiello and Dunbar, 1993, 191). This may be supported by the fact that these primates are primarily found in forest-edge or woodlands, while early Homo likely occupied more open, savannah grasslands. Furthermore, Gelada babboons that live in grassland habitats can have populations of up to 270 individuals (Dunbar, 1984). The second possibility for increased group size is for predator protection, and protection from other Homo groups. This suggestion is found in Darwin (1871a). Increased body size may also be as a response to increased intraspecific competition (Aiello and Dunbar, 1993, 191). A third possibility suggests that, "given that large-scale migration seems to be a major human adaptation, it may have been necessary to ensure that migrating groups had ready access to water holes and food sources over a very wide area" (Aiello and Dunbar, 1993, 191). This view is supported by modern hunter-gather groups, which consist of smaller groups which split from the larger group to hunt. Unfortunately, as Aiello and Dunbar (1993, 191) lament, "this problem appears to be insoluble". However, what is clear is that this interaction of social group and encephalisation is key, and that the evolution of larger groups and associated language evolution "need not have had the same cause or combination of causes throughout their course" (Aiello and Dunbar, 1993, 191).

#### 3.3.1.2 HOMO ERECTUS AND MUSICAL SIGNALLING

As we have seen, Homo erectus was adapted to hunt in the savannah, using sophisticated stone tools and axes. Thanks to this increased meat in the diet, encephalisation is theorised to have occured at this stage. These points of evidence support the conclusion that the proposed musical signalling system may have emerged with Homo erectus. The aim of this discussion is to locate the earliest possible point at which this musical signalling system may have emerged. Homo erectus has not been proven to be ancestral to Homo sapiens, however it is proposed that Homo sapiens had the same prerequisite abilities as seen in Homo erectus. In fact, recent analysis "points to Homo ergaster, rather than Homo erectus, as the direct ancestor of Homo sapiens" (MacWhinney, 2002, 245). However, as stated by MacWhinney (2002), the relationship between Homo erectus and Homo ergaster is not clearly understood.

Aside from fossil evidence of the *outcome* of encephalisation, we can look at endocasts of Homo erectus brains, and make tentative conclusions about its ability to create vocal signals. Wynn (1998), in a study of endocasts, found that the anatomy of Broca's and Wernicke's areas in the Homo erectus brain suggests that they are human-like. However, the anatomy of the left parietal lobe and lower left frontal lobe "include, but are not coextensive with, Broca's and Wernicke's areas... we can conclude that these general areas of the Homo erectus brain are modern in shape, but must be more cautious about identifying the presence of speech production and comprehension areas" Wynn (1998, 78). Furthermore, as the precise role of both areas is not yet clear, it would be hasty to attribute full language abilities to Homo erectus.

MacLarnon and Hewitt (1999), in a study of the thoracic vertebra in non-humans, adult humans, and fossil hominins found that "the thoracic vertebral canal of early fossil hominids was of similar relative size to that of extant nonhuman primates, and substantially smaller than that of modern humans" (MacLarnon and Hewitt, 1999, 347). Finding that the level of thoracic ennervation in Neanderthals is similar to modern humans, the authors conclude that "enhanced breath control, which is a necessary feature for fully modern language, therefore was not possible for earlier hominins up until at least 1.6 mya, the time of *Homo ergaster* (or early *Homo erectus*)". On the basis of this information we might tentatively conclude that Homo erectus did have at least the breath control necessary for making sequences of complex musical signals. Further evidence supporting this approach comes from Deacon (1997), who states that, as Homo erectus had a relative brain size intermediate between modern apes and modern humans, it could plausibly have had "an intermediate level of cortical control over respiration, supported in part by an increase in cortical projections to thoracic motor neurons as well as to other higher respiratory centers" (Deacon, 1997, 252). However, Deacon agrees with Wynn (1998) and MacLarnon and Hewitt (1999) in stating that *enhanced* vocal control necessary for speech was impossible for Homo erectus, claiming that although the speech of Homo erectus "would not have had either the speed, range, or flexibility of today…it would have at least possessed many of the consonantal features also found in modern speech" (Deacon, 1997, 358). However, Capasso et al. (2008), in discussing a hyoid bone attributed to Homo erectus (dated to 400,000 years ago), has concluded that "the shield-shaped body, the probably small size of the greater horns and the radiographic image appear to be archaic characteristics…suggesting that the morphological basis for human speech didn't arise in Homo erectus" (Capasso et al., 2008, 1007).

As we are not here positing the emergence of speech or language with Homo erectus, this argument does not pose problems. If we conclude that Homo erectus had some voluntary breath control, aided by cortical control, alongside the ability to make vowel sounds, this is enough to support the view that Homo erectus may have been the first Homo species to exhibit and use a musical learned sequential precursor. As mentioned earlier, this precursor is proposed to have been multimodal in nature, however, evidence of use of instruments or bodies to create musical sounds has not fossilized.

MacWhinney (2002, 235) proposes three periods in the evolution of language. In the first, from eight to four million years ago, he states that evolution focused on "the linking of morphological and neurological changes to the attainment of bipedal gait". The second period, from 4 to 2 million years ago, saw "the development of methods for solidifying group structure during face-to-face vocal interactions" (MacWhinney, 2002, 235). It is with the third period, from two million years to 100,000 years ago, that “the focus of co-evolution was the linkage of a broad set of symbolic processes to neural control mechanisms. These linkages involve the growth of prosody, chant gesture, dance, and a variety of other largely social expressions of language functioning” (MacWhinney, 2002, 235). Whilst some of MacWhinney’s dates seem a little early, I agree with his discussion of the emergence of prosody, chant, gesture and dance, with Homo erectus. As MacWhinney (2002, 245) writes, “during this period [before two million years ago] the species had achieved some level of group solidarity through the social use of vocalisation…the first groups of Homo erectus were then able to elaborate new forms of symbolic communication in both vocal and gestural modalities”. Whilst the proposed musical learned sequential precursor does not posit a *symbolic* communication system, I agree that a previously established solidarity or cohesion system based on vocalisations could have been co-opted for signalling of fitness.

Dunbar (2003), in an attempt to cast light upon the timing of the evolution of language, has drawn upon work from Aiello and Dunbar (1993) relating social grooming time to group size in primates. Dunbar’s theory is that, once group size moved past a size which allowed individuals to physically groom each other (as we see in other primates), a system of vocal grooming (language) must have evolved. An individual in a rapidly expanding social group simply did not have the time to groom other individuals to establish and maintain social cohesion, and so a pressure for a physically distant system, with which one could groom multiple hearers, emerged. Dunbar (2003) took the brain volume measurements of hominins, and regressed from that the proportion of time which must have been spent grooming in order to have a cohesive social group. He observes that catarrhine primates have a maximum limit of grooming time at 20%, and allows a further 10% for vocal grooming and selection pressures, giving an upper limit of 30% of the day spent grooming, before the pressure for an alternative forced the evolution of language as social grooming. This provides a rubicon of 30% “beyond which group size could not increase unless language came into play” (Dunbar, 2003, 174). This regression technique suggests that language “at least in some form”, must have evolved by around 0.5 million years ago. Further, “the distribution of data-points on the graph suggests that Homo erectus populations, taken as a whole, simply sit astride this rubicon, but Homo sapiens populations all exceed it” (Dunbar, 2003, 175). Dunbar takes this as evidence that Homo erectus “almost certainly lacked language capacity” (Dunbar, 2003, 175), as the mean predicted grooming time required for Homo erectus to maintain social bonds was 30.97% (Aiello and Dunbar, 1993, 190). It is only with Homo sapiens that early hominins would have undergone the same pressure on social bonding as that found in modern human societies.

Aiello and Dunbar (1993) characterise four stages in the evolution of language, beginning with 1) a conventional primate grooming based process, found in australopiths. This was followed by 2) the increased use of vocal chorusing to bond groups, which was characteristic of Homo erectus. The third process emerged with the appearance of a socially focused language, which expanded the range and quality of interactions supporting larger groups. This stage was associated with the appearance of Homo sapiens. These three initial stages all facilitated the fourth stage language as we know it now.

Dunbar (2003) states that "a plausible interpretation" of stage 2 (a phase of increasingly extended vocal exchanges) is that "language evolved out of wordless singing . . . and. . . music had an early (and perhaps separate) development as part of the bonding mechanism that welds human groups together" (Dunbar, 2003, 176). This "plausible interpretation" is heartening to see, and provides support for the ever-growing literature into music's role in the evolution of language. However, I would suggest a re-analysis of stage 2 as consisting of a musical signalling system which provided cues of fitness. This stage would have facilitated the third stage, of a socially focused language, by scaffolding a communication system that was based on social fitness cues and interactions, which underwent selection, driven by collaborative partner choice.

Dunbar outlines two possible pathways for language evolution based upon this implication. The first suggests that "music has a very ancient history, dating to the early Homo erectus period" (Dunbar, 2003, 177). Under this scenario, grammatical language evolved with the appearance of archaic humans approximately 0.5 million years ago. The alternative hypothesis is that "the sea change in grooming time requirements at 0.5 MY. . . marks the point at which musical chorusing evolved into a particularly intense form" (Dunbar, 2003, 177), with the musical interactions before this, during the Homo erectus period, resembling counter-calling as seen in modern primates. Grammatical language would then have been a later development, perhaps with modern humans around 200,000 years ago. Dunbar himself argues for the first hypothesis, writing that "the necessity for both large groups and (at least rudimentary) language appeared early in the evolution of the genus Homo and began to increase rapidly in the second half of the Middle Pleistocene" (Aiello and Dunbar, 1993, 184). There was a "gradual outgrowth of the evolution of enhanced vocal communication that first got under way almost 2 million years ago" (Aiello and Dunbar, 1993, 190). This thesis cannot hope to address all questions of the evolution of language, but with regard to this point, we can suggest that music has, as Dunbar put it "a very ancient history", and date the emergence of the LSP to the time of Homo erectus.

## 3.4 Discussion

This chapter has proposed the evolution of language from the learned musical precursor outlined in chapter two. This proposed precursor acted as an honest signal of cognitive ability, and was made up of complex sequences of rhythmic, musical sounds. Honesty was maintained by the investment in learning required for such a system to be learned, as well as the cognitive abilities necessary to learn the system. This system was an ever-changing, culturally learned system, which maintained honesty. This hypothesis is richly supported by analogy with birdsong and developmental stress research. This musical signalling system is proposed to have been in place from approximately 1.6 million years ago, with the advent of Homo erectus, a near-modern species of Homo which is hypothesised to have had increased brain capacity and sophisticated cultural artefacts such as stone tools. Musical signalling would have provided cues for perceivers, which allowed them to evaluate the fitness of a producer. This system is proposed to have undergone social selection driven by collaborative partner choice, as both perceiver and producer needed to choose a partner for collaborative endeavours. Adopted by Homo sapiens, this culturally learned system was passed from generation to generation, and pressures operated on it, as we have seen in both modelling and experimental work, breaking down what was a holistic signal to a sequence of sounds - "proto-words" which became associated with meanings. This may have been the beginning of a meaningful system - a protolanguage. The next chapter will outline the beginnings of an experimental paradigm which aims to test the predictions made in this chapter. Chapter five provides further experimental investigation of the proposed selection mechanism that the LSP may have undergone, social selection.

# CHAPTER 4

# EMPIRICAL TEST OF HYPOTHESIS

Research into musical predecessors to language has, to date, been hampered by a relative lack of empirical research. While musicology has a strong, rich empirical background, and empirical research into the origins and evolution of language is growing apace, it is only in recent years that an attempt has been made to test hypotheses about language evolution from a musical predecessor. It should be noted that work by, for example, Honda and Okanoya (1999) and Okanoya (2004) has provided a rich basis for experiment, and work by Fitch and Rosenfeld (2007) and Charlton et al. (2012) adds to the growing empirical literature. There are obvious reasons why musical protolanguage theories may appear impossible to test - for one, vocalisations and early musical behaviours do not fossilise. Secondly, modern humans, and societies, do differ from the population under question in theories of language and music evolution. More practical considerations exist, such as the difficulty of defining exactly what "musical behaviour" amounted to in hominin evolution. Furthermore, cynicism from some quarters, for example from Weiss (1974, 104), who rejects the "speculation that mankind began its linguistic career by singing like the birds" as "pure quackery", and a lack of creative thinking has hampered research in this area. However, we cannot let these limitations restrict us from generating testable hypotheses, and testing them. In particular, Tecumseh Fitch (2006a, 2010, 2012) has written of the need for empirical investigation into musical predecessors of language. It is in this spirit that the following two chapters present empirical investigation into the hypothesised musical signalling precursor to protolanguage.

In this chapter I will present two studies, in an attempt to ground the hypothesised LSP in experimental data. The first study presented here explores the role of complexity in perceivers' perceptions of the quality of a signaller. I propose that signallers could have used complex sequences of musical sounds for the honest signalling of cognitive ability. Thus, we might expect perceivers to have a preference for complex over non-complex rhythms. The second study casts light upon the proposed relationships between musical signalling and cognitive ability.

## 4.1 EXPERIMENTAL ANALYSIS OF COMPLEXITY

As seen in chapter two, the hypothesis outlined in this thesis suggests that complex sequences of musical sounds could have acted as signals of cognitive ability in an early hominin society. In order to further investigate this hypothesis, complexity and its role in communication should be analysed. A study was conducted which aimed to identify any underlying human preferences for complexity, and additionally to identify whether increased complexity was associated with other traits. This involved a rating paradigm which will be outlined below, following a discussion of complexity as a testable phenomenon.

There is a multitude of views of the meaning of the term "complexity". In general use, people seem to take this term to mean something like "many items, arranged in an intricate or multi-faceted manner". The study of complexity has suffered from the attitude exemplified by the saying "a scientist would rather use someone else's toothbrush than another scientist's nomenclature" (Gell-Mann, 2003, 51). Indeed, there are so many disciplines concerned with complexity – computational complexity, ecological complexity, economic complexity, political complexity, social complexity (Erdi, 2008), that a consensus on the use of terminology would be surprising. Often real-world examples, such as ant colonies, brains, and the immune system, are used to exemplify complexity, without explicitly addressing the meaning of the term.

Conley (1981, 59) has written, "the term complexity has not been adequately defined nor has a satisfactory measure of it been established". This notion has been studied extensively in the context of information theory, and is "deeply connected" with concepts such as randomness, information, regularity and coding (Schmulevich and Povel, 2000, 61). Schmulevich and Povel further define complexity according to the amount of information embedded within an object (in our case, a sequence of sounds). Furthermore, complexity measures can be obtained by asking participants to judge stimuli according to how difficult it would be to learn and reproduce.

Early experiments (Berlyne, 1958) showed that participants fixated for a longer period of time on complex rather than non-complex figures, "the tendency to spend more time on more complex figures was significant beyond the .01 level with every category" (Berlyne, 1958, 293). Here, Berlyne measured complexity by the amount of irregularity of material in the figure, the heterogeneity of elements, the irregularity of shape, incongruity, and incongruous juxtaposition. Additionally, Fantz (1961) has reported studies which show that infants attend to complex patterns more than non-complex ones. So, humans seem to show a preference for complexity. The following section will outline potential measures of *musical* complexity.

### 4.1.1 Measures of musical complexity

Music can be seen as a struggle between the need to be innovative and create something new and exciting, and the need to remain within the pre-established boundaries of a musical paradigm; novel but cultural. Few great composers and musicians have been pure innovators - instead, they have largely built upon the strong basis developed by their predecessors whilst adding some novel elements. It is only the rare few who attempted to entirely revolutionize the musical world - for example Arnold Schoenberg, with his invention of the twelve-tone or serial system (although Schoenberg himself stated "I was never revolutionary. The only revolutionary in our time was Strauss!" (Schoenberg, 1975, 137)). Indeed, despite efforts by Schoenberg's students Anton Webern, Alban Berg, and Hans Eisler, the twelve-tone or serial approach to composition never took precedence over the established tonal system. So, there must be a pressure both to innovate, which adds interest, novelty and expressivity, and also a pressure to retain values that mark a composition as part of a pre-determined system. Thus we have a tension between orthodoxy and novelty. Novelty, expressivity and diversity can be introduced with (rhythmic, melodic and harmonic) complexity.

Many factors contribute to the judgement of musical complexity. According to Finnas (1989, 6) these can include "unusual harmonies and timbres, irregular tempi and rhythms, unexpected tone sequences and variations in volume". Some measures of complexity employed in empirical studies include the number of different pitches, melodic redundancy (degree of randomness) (McMullen, 1974), syncopation and percentage of minor and major chords (Heyduk, 1975), syncopation (Longuet-Higgins and Lee, 1984; Fitch and Rosenfeld, 2007) rate of rhythmic activity (Conley, 1981), and number of tones/ bits of information per second (Vitz, 1964). Previous studies (Smith and Melara, 1990; Beauvois, 2007; Heyduk, 1975; Steck and Machotka, 1975;

North and Hargreaves, 1995; Ziv and Keydar, 2009) have investigated the relationships between musical complexity and preferences; however, the following study aims to add the judgement of cognitive and physical fitness to standard questions about preference and familiarity.

In a 1981 study into complexity in music, Conley (1981, 60) sought to pinpoint the relationships between "physically specified variables" in music and judged complexity. In order to investigate this, a list of ten measurements of complexity was made, including the regularity or irregularity of numbers of tones in vertical structures, the number of independent parts, number of different harmonies, number of changes in harmony, number of measures of each of tonic and dominant harmony, number of changes of other harmony, number of changes of rhythmic activity, rate of rhythmic activity, and duration. Three groups of students - graduate music students, undergraduate music students, and undergraduate non-music students - were asked to rate the complexity of pieces (Beethoven's Eroica variations) which differed according to the above measurements of complexity. It was found that nine specified music variables (items six and seven, number of changes of dominant harmony and of other harmony, were conflated due to high correlation) accounted for 95% of the variability of all participants' ratings. Each group's judgements of complexity were highly correlated, regardless of differing musical background.

For all groups, the correlation of the complexity scores with item nine, rate of rhythmic activity, was significant ($p = .001$). Stepwise regression showed that this variable alone could account for 71%, 90%, and 78% respectively of the variability of responses from graduate music students, undergraduate music students, and undergraduate non-music students. This was the best one-variable model for all three groups. The best two-variable model was rate of rhythmic activity and changes in rhythmic activity, and the best three-variable model was the above two and number of measures of other harmony. This indicates a good starting point from which to approach complexity, that of rate of rhythmic activity. As Conley states,

> "The variable X9 (rate of rhythmic activity) was a powerful measure of complexity, having extremely high correlations with the judgements of all groups and accounting for a huge percentage of the variability in the responses of all groups. In all cases X9 accounted for more than twice the variability of that accounted for by the next most highly rated variable" (Conley, 1981, 69).

Conley tentatively states that the experimental paradigm employed in the article may enable us "to determine a meaning of a term or concept [complexity] used in describing music and music experience that had previously been undefined or only vaguely defined" (Conley, 1981, 70). However, there may be some flaws in the experimental paradigm. The participants in this experiment were asked explicitly to rate the complexity of the pieces. However, we do not know what preconceived notions people are bringing to a task such as this. There is a myriad of ideas of what constitutes complexity, and participants may be bringing any number of these to the experiment. A more implicit task may better approach this question.

In a 2008 study into complexity in birdsong, Boogert et al. (2008) found that total number of song phase elements in birdsong was a significant predictor of learning performance, with males whose song phases contained more elements requiring fewer trials to solve the final level in a novel foraging task. Thus, total number of rhythmic elements may be a worthwhile avenue of investigation as a working definition of complexity. Boogert et al. (2008) measured "total number of song phase elements" by analyzing the waveform amplitude. When the waveform returned to baseline, adjacent elements were treated as separate. This can be interpreted as total number of song events. Although this was based on analysis of birdsong, I feel it can adequately be applied to a rhythmic paradigm such as ours.

Conley (1981, 62) defined "rate of rhythmic activity", the measure of complexity which correlated most highly with rated complexity of musical pieces, as "the basic, collective number of rhythmic events per minute" which was calculated by setting a metronome to coincide with the "basic movement…the smallest constant note value of the collective rhythm". I interpret this to mean the number of rhythmic events (beats) per minute. Note, this is not the meter, the underlying rhythmic pulse, but rather, the beats of the overlying rhythm.

Thus, we can combine these two similar measurements of complexity to suggest a new measure to be used in our experimental paradigm. That is, complex rhythms will be those with *most* rhythmic elements. So, a piece with a simple 4/4, crotchet-based rhythm with minimal elements will be judged non-complex, and a piece with a 4/4 rhythm based on quavers, semi-quavers and so on, will be judged complex. It should be noted that the measurement of complexity here resembles that used by Charlton et al. (2012). In that paper, complexity was introduced by differing note duration and syncopation. In the present study, we introduce different note duration, which was found by Charlton et al. (2012, 4) to have a significant effect on complexity ratings: "increasing note durations had a much greater effect on women's perceived complexity judgements than increasing pitch range". The findings of the Charlton et al. (2012) paper supports the measure of complexity used here.

### 4.1.2 EXPERIMENTAL TEST OF COMPLEXITY HYPOTHESIS (STUDY ONE)

A central question of the proposed study is whether complex sequences of musical elements can or do signal cognitive ability. In order to answer this question, we must investigate whether perceivers really do have a preference for complexity in musical pieces. This study aimed to examine whether participants responded to complex and non-complex pieces of music differently, given no training in either. Participants listened to examples of both complex and non-complex pieces of music (each eight seconds long). These rhythms can be seen in western notation in Figures 4.1 and 4.2. They were told “You are an explorer, who has just discovered an isolated human society, who have had no interaction with the outside world. You will now hear some examples of the drumming these people make. Your job is to answer questions about the rhythm and the person who made it”. Participants were played four complex and four non-complex rhythms, and asked to use a seven-point Likert Scale, with poles of “not at all” to “very/very much”, to answer five questions about each rhythm;

1. How much do you like this rhythm?
2. How familiar is this rhythm?
3. How likely are you to pick this person to do a tricky treasure hunt with? (“collaboration” question)
4. How attractive do you think this musician is?
5. How intelligent do you think this person is?

Full instructions given to the participants can be found in the Appendix, on page 159. Participants were asked about the familiarity of the piece to test whether familiarity with stimuli is necessary to make judgements of quality, preference, and collaboration likelihood. Further, this might indicate whether non-complex or complex rhythms are seen as more “familiar”, even though both are entirely new to the participants. It was hypothesized that complexity of the stimulus would correlate positively with personal preference judgements, collaborative likelihood,

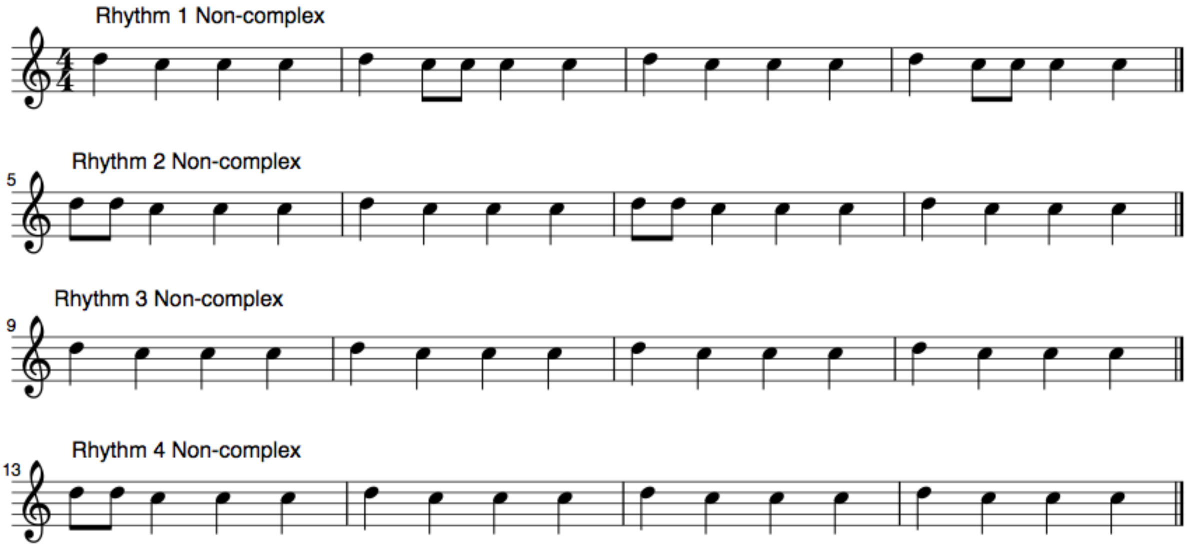



Figure 4.1: Non-complex rhythms used in study one

and rated attractiveness and intelligence. Participants were prompted to answer each question with their first instinct.

Fifteen participants (F = 10, M = 5), all PhD students in the University of Edinburgh, underwent this study. Experiments were run on an Apple MacBook running OSX Leopard, using a Psyscope script. Psyscope, developed at Carnegie Mellon University on the 1990s, is designed to run psychological experiments, and is widely used in both linguistic and psychological studies in the University of Edinburgh. This script presented each of eight rhythms, matched with each of five questions (seen above). Participants answered the questions by selecting a key numbered between 1 and 7. Participants were given Sennheiser headphones to cut out environmental noise and to present the rhythmic stimuli. They were given vocal and written instructions, and told that the experimenter was close by to answer any queries during the study. Experiments took on average eleven minutes to run. Stimuli rhythms were composed by the author and can be seen in Figures 4.1 and 4.2. These compositions were synthesized to create auditory stimuli, using the Ableton project synthesizer. The resulting auditory files were each eight seconds long, and consisted of two woodblock sounds. This study was approved by the University of Edinburgh linguistics department ethics forum. All participants signed consent forms to take part in the study, and for their data to be stored anonymously on a secure server.

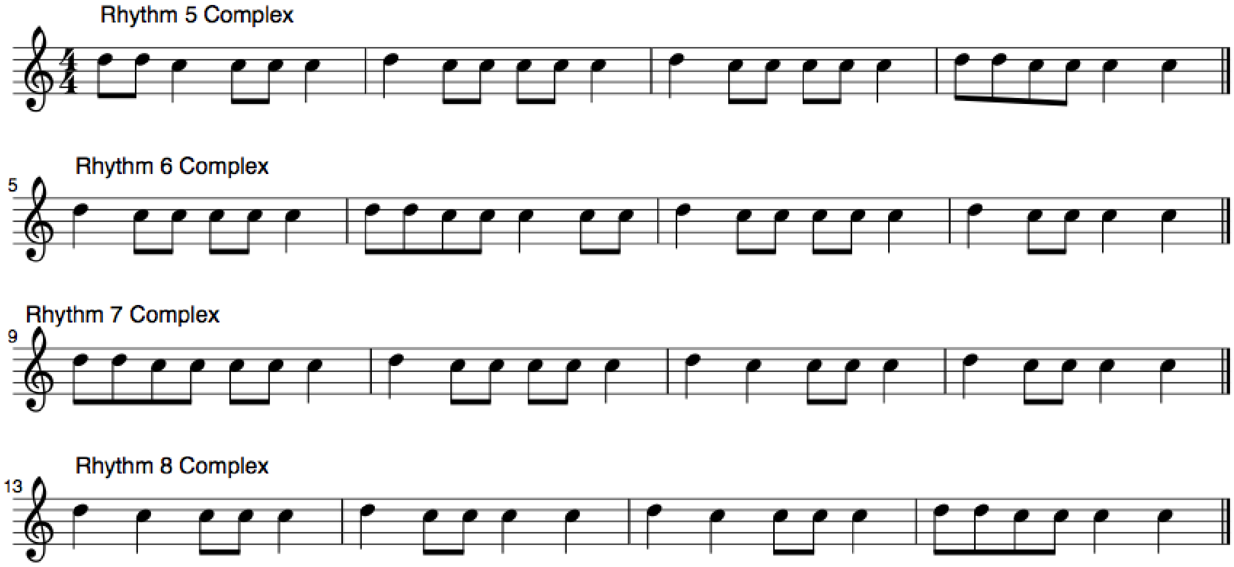



Figure 4.2: Complex rhythms used in study one

#### 4.1.2.1 RESULTS

As can be seen in Table 4.1, the ranges and mean scores for complex and non-complex rhythms differed, with complex rhythms being rated higher than non-complex, except for in the familiarity condition.

| | Complex | | Non-Complex | |
|---|---|---|---|---|
| | Range | mean | Range | mean |
| Like | 3.00 - 6.75 | 4.84 | 1.50 - 6.50 | 3.55 |
| Familiar | 2.00 - 6.25 | 4.41 | 4.25 - 6.25 | 5.43 |
| Collaboration | 2.75 - 6.75 | 4.98 | 1.50 - 6.75 | 3.64 |
| Attractiveness | 2.25 - 6.50 | 4.96 | 1.50 - 6.50 | 3.52 |
| Intelligence | 2.50 - 7.00 | 5.11 | 1.25 - 6.75 | 3.46 |

Table 4.1: Use of scale and mean values in study one

Kolmogorov-Smirnov tests revealed that the data was non-normally distributed, and so non-parametric tests were used in this analysis. Spearmans Rho correlations were conducted on a 2-tailed hypothesis, that complexity would have some effect on the ratings for each question. Due to the number of comparisons and the related risk of type 1 errors, we performed a Bonferonni correction on the alpha, correcting it to 0.01. Significant correlations were found between complexity and preference (like) ($r_s$(26)=.50, p=.007), attractiveness ($r_s$(26)=.55, p=.002), and intelligence ($r_s$(26)=.54, p=.003). See Table 4.2.

These positive correlations indicate that participants preferred complexity in rhythms, and found more attractive and intelligent, “musicians” who created complex pieces. One of the important variables of interest, intelligence, was significantly correlated with complexity. A trend towards a negative correlation was found between complexity and familiarity, ($r_s$(26)=-.43, p=.022). This means that participants tended towards rating as more familiar the non-complex rhythms they were presented with. This is not surprising, as participants possibly found more predictable, simple rhythms more familiar as they reflect simple rhythms we hear every day - the beep of an alarm clock, for example. The more complex rhythms add an element of unpredictability, surprise, and so, were scored as less familiar than the complex ones.

The interaction of all variables is worth investigating. As can be seen in Table 4.2, preference (“like”) correlated significantly with cooperation, attractiveness, intelligence, and complexity. It did not correlate with familiarity. Cooperation correlated with attractiveness and intelligence, and attractiveness correlated with intelligence and complexity. So, we see that the only variable not to correlate with the other variables is familiarity. The level of inter-correlation seen here supports the use of a Bonferonni correction of the alpha.

| | Like | Familiar | Cooperation | Attractiveness | Intelligence |
|---|---|---|---|---|---|
| Familiar | .09 | – | – | – | – |
| Cooperation | .85** | -.11 | – | – | – |
| Attractiveness | .86** | -.01 | .76** | – | – |
| Intelligence | .81** | -.06 | .76** | .92** | – |
| Complexity | .50** | -.43 | .48 | .55** | .54** |

Table 4.2: Correlations between complexity and other variables
*Note*: ** indicates that correlation is significant at 0.01, the corrected alpha value

Whilst we have seen significant correlations between complexity and other variables, a Wilcoxon matched paired Signed Rank Test was conducted to investigate whether the difference in scores for complexity and non-complexity was significant. The Wilcoxon test showed that the difference between complex and non-complex stimuli was significant for all judgements (Table 4.3).

|  | NonComplex |  | Complex |  | Significance |
|---|---|---|---|---|---|
| Variable | mean | s.d. | mean | s.d. | p value |
| Like | 3.55 | 1.25 | 4.82 | 1.19 | p=.003** |
| Familiar | 5.42 | 0.79 | 4.41 | 1.35 | p=.011* |
| Cooperation | 3.64 | 1.16 | 4.98 | 1.16 | p=.028* |
| Attractive | 3.52 | 1.18 | 4.96 | 1.28 | p=.002** |
| Intelligence | 3.46 | 1.39 | 5.11 | 1.35 | p=.001** |

Table 4.3: Difference between complex and non-complex stimuli
Complex rhythms were rated higher for all questions but familiarity. p values show the significance of difference in rating of complex and non-complex stimuli. ** indicates that the difference is significant at 0.01, * indicates that the difference is significant at 0.05

#### 4.1.2.2 Discussion

Using the new measure of complexity (number of rhythmic elements) we found a significant correlation between complexity and all rated measurement questions. Raters preferred more complex pieces, preferred to cooperate with "musicians" (in reality, all of these rhythms were synthesised) who created more complex rhythms, and found these "musicians" more attractive and intelligent. The correlation between complexity and familiarity was in a negative direction, but not significant once the Bonferonni correction was applied. This trend is easily explained by the unexpected, surprising nature of the complex rhythms.

The Wilcoxon Signed-Rank test reported a significant effect of complexity on scores, for all question variables. A central question of the present study was whether complexity is favoured over non-complexity as a signal of cognitive or collaborative skill. Crucially, participants found "musicians" who created complex rhythms more intelligent, supporting the proposal of complex rhythms as a signal of fitness.

It should be noted at this point, that the ratings of the five questions may have influenced one another. As all five questions deal with the attributes of an individual and his/her output, we may have an example here of the "Halo Effect" (Thorndike, 1920; Dion et al., 1972) in which our raters may have become cognitively biased to judge all questions according to the same positive approach. In other words, if the complexity of the piece led the rater to see the participants as "better", then this might effect ratings of attraction, intelligence etc. However, one counter-example to this view exists in the lack of correlation between complexity and familiarity. If complexity were associated with a general "positive halo", then we might expect familiarity to positively correlate with complexity. Even if this Halo Effect is in place, it should not invalidate our findings, as we already know that each variable, excluding familiarity, is highly intercorrelated (Table 4.2).

At this stage, it appears that our prediction is upheld, that participants would prefer complexity across a number of measurements. This supports the hypothesis that complexity may have played a role in signalling cognitive ability in a learned musical system.

These results show no evidence that collaborative partner choice *alone* is in operation - in other words, the ratings for preference and attractiveness judgements suggest that mate choice may also have been an important driver of selection. Indeed, there seems no need for the two hypotheses to be mutually exclusive. Chapter five will further address the question of selection and choice. The next section presents a study testing potential relationships between musical learning ability and cognitive ability.

## 4.2 NON-SEXUAL SELECTION, MUSICAL LEARNING AND COGNITIVE ABILITIES

One aim of this thesis is to tease apart the roles of sexual selection and other selection mechanisms. As can be seen in chapter one, the prevailing view on the selection of musical precursors to language remains that of sexual selection. However, this thesis champions *social* selection, defined here as differential reproductive success as a result of differential success at social interactions - this is in contrast to sexual selection, defined here as differential reproductive success as a result of success and finding mates. This was discussed in more detail in section 2.5 on page 56. Studies two and three outlined in this thesis aim to directly address the possible selection mechanism for the LSP.

One way to address this disparity between sexual and social explanations is to generate predictions about the relationships we might find under either scenario. The proposed learned musical signalling system served as a signal of cognitive ability. Thus, we should expect to find correlations between cognitive abilities and musical learning abilities. Such correlations would support the view that the LSP acted as an honest signal of cognitive ability, allowing perceivers of these signals to make informed collaborative partner choices. If this system were assumed to have undergone *sexual* selection (the prevailing view in the field), then we might expect to find further relationships between cognitive abilities, musical learning abilities, and direct measurements of physical quality (these measurements presumably being genetically underpinned). If, however, this system is assumed to have undergone *social* selection (driven by collaborative partner choice), then we might expect no correlations between measurements of physical quality and musical and cognitive abilities. In a social selection scenario, in which reproductive success is a function of social success, there is no choice on the basis of genetics, and so physical quality, and good genetics should be irrelevant.

Developmental stability, a measure of a body's ability to develop well despite perturbation, can be seen as a measure of physical quality. This measure can be indicated by both cognitive ability and fluctuating asymmetry (a measure of deviation from what is symmetrical at a population level). This is a property of an individual which is hypothesised to reflect the resistance of this individual to stresses in development. If no correlation exists between fluctuating asymmetry and musical abilities and underlying cognitive abilities, then there would be no pressure for musical output to undergo sexual selection. This supports the inclusion of social selection as a potential important selection mechanism. The following study, and the study outlined in chapter five, aim to highlight social selection as potentially very important to the evolution of the LSP.

Fluctuating asymmetry is proposed as an inverse index of developmental stability. Fluctuating asymmetry (FA) is a reliable indicator of developmental instability, and is an empirically measurable phenomenon. There is already a large literature highlighting links between FA and intelligence, and so FA seems an appropriate basis for a study testing the quality of musical sequences as a signal of developmental stability (if undergoing sexual selection). Cognitive ability will be measured using two independent measures (processing speed and general intelligence), and the relationships between fluctuating asymmetry and cognitive ability will be examined. Additionally, the ability of subjects to learn to use a novel musical instrument will be tested (as a test of ability to learn a complex, novel musical system), and the results of this test correlated with the cognitive ability and symmetry data. This is the first step in a hopefully rich and fruitful empirical approach to determining musical predecessors to language.

### 4.2.1 FLUCTUATING ASYMMETRY

Fluctuating asymmetry (FA) is a measure of an individual organism's deviation from what is symmetrical at the population level, as measured bilaterally. It is defined as "random deviation from perfect bilateral symmetry in a morphological trait for which differences between the right and left sides have a mean of zero and are normally distributed" (Watson and Thornhill, 1994, 21). It can be measured using finger lengths, by measuring a certain organ of the body, or, to be as accurate as possible, by a combination of both of these approaches. The standard method of computing FA is by taking two independent measurements of some parts of the body - i.e. the foot, ankle, wrist, elbow, ear breadth, and ear length and third, fourth and fifth finger lengths. Two measurements of each are taken with a digital calipers in order to reduce experimenter error, and these are then averaged to provide an accurate measure of each part of the body, on both sides. If participants have undergone any trauma to any of the body parts under study, e.g. broken wrist, these data are usually not used. Then the absolute percent asymmetry for each body part is calculated using the following formula, which is the standard formula for FA, and has been used in most research to date (Bates, 2007; Furlow et al., 1997; Prokosch et al., 2005). These absolute percentage values are summed to provide an overall FA score.

$$\Sigma(|(\text{left-right})/(\text{left+right})/2|)$$

This formula refers to the sum of (all measurements of the left hand side subtracting the same measurements of the right hand side, divided by the sum of left and right measurements, all divided by 2). In recent years, 3D scanning for symmetry has emerged (Brown et al., 2008; Claes et al., 2011). These scanners provide standardised, automated, quick measurements, providing many more measuring points and repeated measures than could be reliably measured by an experimenter.

Similar to birdsong, FA operates as an reliable indicator of developmental stability or resilience – the more symmetrical an organism, the better resistance it has had to developmental instability (stresses). Both sides of an organism can be seen as independent replicas of the same developmental event. They share the same genotype, and, in a homogenous environment, are under the same environmental influences. However, during development, small random perturbations act upon the developmental pathway of the organism, causing it to deviate from the expected trajectory. These processes can act locally, on only one body part, and lead to asymmetrical body types (Dongen, 2006). So, the ability of an organism to develop normally despite these perturbations is a measure of its resilience.

Research has identified associations between FA and a variety of human traits, for example social dominance (Furlow et al., 1998), cognitive aging (Penke et al., 2009), working memory (Yeo et al., 2000), and intelligence (Furlow et al., 1997; Bates, 2007; Prokosch et al., 2005). Various studies into the roles of fluctuating asymmetry and intelligence have identified a negative correlation between FA and IQ, as measured by the Raven Standard Progressive Matrices (Bates, 2007), Raven Advanced Progressive Matrices (Prokosch et al., 2005) and Cattell's Culture Fair Intelligence test (Furlow et al., 1997). Using two independent samples, Bates (2007) found significant correlations between FA and IQ of -0.41 and -0.29, and concluded that mental ability is a reflection of general fitness factors which influence the development of all body systems, "with brain function being an especially sensitive indicator of this fitness factor" (Bates, 2007, 41). An earlier study by Prokosch et al. (2005) had studied men only, and sought to identify which of five tests would best predict FA. It was found that the most g-loaded (i.e. general mental ability) test was the advanced Raven's Progressive Matrices, and that "there is a significant association between the extent to which a cognitive test indexes morphodevelopmental stability…and…neurodevelopmental stability" (Prokosch et al., 2005, 206). Furlow et al (1997) was one of the earlier studies to correlate IQ and FA, and they found that, in two samples, "FA correlated negatively and significantly with CFIT [intelligence measure] scores" (Furlow et al., 1997, 823).

Using multiple meta-analytic methods, Banks et al. (2010) studied the relationship between FA and g (general mental ability) in a number of studies, and found that "the correlation between FA and g is negative with estimates of the population correlation ranging from -.12 to -.20" (Banks et al., 2010, 399). They conclude that "the relationship between g and fluctuating asymmetry is not large, which indicates that common and distinct genetic and environmental factors likely contribute to variance in both …the relationship between FA and g may be explained by their association with a general fitness factor". A further meta-analysis, conducted by VanDongen and Gangestad (2011), used different statistical methods and theoretical assumptions, and also found a significant relationship between FA and g, with an effect size of 0.11, from ten sampled studies with a total N of 1071. This supports the view that the relationship between FA and g is a reliable, substantial effect.

The outcome of these studies, as summarised by Banks et al. (2010), has been a consensus that the significant negative correlation between general intelligence and FA is as a result of general fitness factors influencing the growth and maintenance of *all* bodily systems, including brain functions. In other words, variation in mental ability is not related to a number of "genes for intelligence", but rather, reflects fitness-related variation across the genome, which is expressed in the body as well as the brain. In this case, individuals who have low FA and high IQ have such because of general fitness - this may be because they were not exposed to lots of developmental stress, or because they proved resistant to the developmental stresses they *were* exposed to – rather than because FA is an inverse signal, per se, of intelligence. The brain is a particularly sensitive indicator of developmental stability, and FA is also indicative of reduced resistance to developmental instability or low metabolic efficiency.

Developmental Stability refers to an organism's ability to develop the appropriate species-specific phenotype, despite perturbations at the genetic and environmental levels, including mutations, parasites, inbreeding, starvation, pathogens or toxins. We can imagine this as analogous to the builder's ability to turn a blueprint (the genotype) into a finished, well-constructed house (the phenotype) (Prokosch et al., 2005). We can view differences in bilateral development as "the outcome of two opposing forces: those tending to disrupt precise development (developmental noise), and those tending to stabilize it (developmental stability)" (Palmer, 1996, 519). The outcome of the interaction of these two forces can create bilateral asymmetry in an individual - fluctuating asymmetry.

According to Prokosch et al. (2005, 203), "just as body symmetry reveals developmental stability at the morphological levels, general intelligence may reveal developmental stability at the levels of brain development and cognitive functioning". Thus, individuals with high Developmental Stability should have low FA and high intelligence. This is supported by the above-mentioned studies into correlations between FA and IQ. Developmental Instability (DI) is a measure of deviations from the intended phenotype (Moller and Swaddle, 1997, 2), and is used to note instability in early development, which can be seen expressed in FA and other traits. DI is assumed to reflect the outcome of developmental noise and a developing system's sensitivity to these perturbations (Dongen, 2006, 1728). So, individuals who had low DI in early development, or who were particularly resistant to such DI, are likely to have low FA and high intelligence, alongside other measures of ability.

The hypothesized musical signalling system was a culturally learned system, which served to signal cognitive ability. If we view DI as an impediment to normal development, we could imagine a situation in which some individuals didn't develop the mental capacity to acquire certain culturally learnt skills, due to a lack of resistance to DI. An individual with high DI - i.e. one prone to genetic and environmental perturbations - may not have developed the requisite brain mechanisms to handle learning to produce a complex multimodal sequence. So, they may have been unable to learn to produce the tricky, complex musical signalling system, whereas one with low DI, or high resistance, and thus high IQ, would have been able to learn such a skill, which would have behaved as an honest signal of their fitness - cognitive or otherwise. Thus, we could plausibly expect a correlation between fluctuating symmetry, our index of DI, in both the social and sexual hypotheses.

We can view Developmental Instability as an analogue of Developmental Stress found in birds (Nowicki et al., 2002; Spencer et al., 2003). Ritchie et al. (2008) sought to model the

Developmental Stress hypothesis in order to investigate two questions: 1) whether a preference for culturally transmitted features of song affected by developmental stress could evolve in a population, and 2) whether cultural transmission of such song features could be stable, despite the inability of members of the population suffering from developmental stress to learn them (Ritchie et al., 2008, 570). In this model, “type 1” songs were learnable by both high and low quality birds while “type 2” songs were only learnable by high quality birds (quality is determined by the level of resistance to developmental stress - a low quality bird is undernourished or parasitised). As the authors write, “it is useful to think of type 2 song as being more complex in some way than type 1 song, e.g. it may have more complex phrase or note structure” (Ritchie et al., 2008, 571). Already here we can see parallels between Developmental Stress and Developmental Instability – we have suggested that individuals with high DI, and low resistance, would not have developed the brain structures necessary for them to learn to produce a complex musical sequence, and so, these individuals would only be able to learn to *perceive* it accurately. Here, “type 2” song could reflect the creative stage we claim is important for musical learning, and relevant to this precursor.

Ritchie et al. (2008, 578) found that “in a range of conditions, selection can favour a genetic learning bias for culturally transmitted song types which reveal a potential mate’s level of developmental stress. . . there can be stable cultural transmission of these less learnable song types”. In other words, a genetic bias for learning culturally transmitted song types may be selected for, as these songs allow female birds to analyse a potential mate’s resilience to developmental stress, and thus, fitness. The authors go on to say that “once an allele encoding this sort of bias is present in a finite population, it will be subject to forces such as drift until some cultural invention or mutation event occurs which produces a song-type with the appropriate features” (Ritchie et al., 2008, 578).

This model seems to support the possibility of the proposed musical sequential signalling system. Birdsong acts as a reliable indicator of developmental stress, as birds with high developmental stress or low resistance to stress are unable to learn to replicate the most complex songs. Similarly, it is proposed that the musical precursor proposed here may have acted as a reliable indicator of developmental instability, as individuals with high developmental instability lacked the brain mechanisms which allowed them to learn to produce complex musical sequences. Crucially, these individuals could likely still learn to *perceive* these sequences, but their lack of *production* skill reliably indicated their developmental instability. As (Ritchie et al., 2008, 578) state: “it is possible that an effect similar to the one we investigate here may have played a role at some stage in the evolution of the human language faculty”.

### 4.2.2 Outline of study two

In order to investigate the connections between fluctuating asymmetry (a measure of resistance to, or simply lack of, developmental instability), cognitive ability (measured here by general intelligence and information processing speed tests), and ability to learn musical signals, a study was conducted, which tested the FA, general intelligence and information processing speed of participants, as well as their ability to learn to replicate and create quality melodo-rhythmic sequences. This study aimed, in part, to shine a light on the sexual vs social selection question. This study was conducted on 104 participants, 52 male. Due to technical failure, the data from two participants, one of each gender, was unusable. Participants fell within the age range of 18 to 37 years, with a mean of 23.03 years, and a s.d. of 3.75. 72.5% of participants were Caucasian, 15.7% Asian, 1% African, with 4.9% of participants identifying as “other”, and 3.9% as “mixed”. There was no information about the ethnicity of 2% of participants. Native language was also noted, with 63.6% of participants naming their first language as English, with 35.3% not English. Data on the first language of 1.1% of participants was missing. Participants had completed a mean of 16.74 years of education, with a s.d. of 2.68, within a range of 6 to 23 years. Participants were recruited from the University of Edinburgh psychology participation research pool, and careers website. Participants received compensation in the form of class credit (psychology students) or £12 for taking part in the entire study, which took on average 1.5 hours.

Fluctuating asymmetry was tested using the $TC^2$ NX-16 3D body and 3DMD facial scanners, newly available to the Psychology Department in the University of Edinburgh. These scanners are accurate to within a millimeter (body: 0.5mm, face: 0.2mm), and lessen the chance of experimenter error. Each scan of a person’s body can be tailored to each experiment’s needs, supplying hundreds of potential measurement points, which can be compared bilaterally. This allows us to calculate an overall FA score for each participant, by using a subset of the possible measurements. Hand scans were taken using a standard flat bed scanner, from which measurements were taken by two independent experimenters.

Additionally, at this stage of the study, general intelligence and information processing speed tests were conducted. These were the Alice Heim 2 tests of verbal, numerical and perceptual ability (Heim et al., 1974a,b) and the Deary-Liewald reaction time test (Deary et al., 2011). This processing speed test consists of two parts, choice and simple. In the simple test, participants must respond, by hitting a key on the keyboard, to a stimulus (X) which appears on the screen. They must respond as quickly as they can. In the choice test, the stimulus may appear in one of four areas of the screen - left-most, left of middle, right of middle, and right-most. Participants must respond by hitting a corresponding key on the keyboard. In total, the experimental battery took on average 1.5 hours. Participants each underwent the same procedure;

1. Greeting, participant information and informed consent
2. Face scanning with 3DMD facial camera system, twice (procedure similar to a seated digital portrait photograph)
3. Body scanning with $TC^2$ NX-16 body scanner, three times (whole scanning takes less than one minute)
4. Scans of both hands using a standard flatbed scanner
5. Questions about former breaks or current sprains/swellings of any of the scanned body parts
6. Deary-Liewald Reaction Time Task (computerized measure of simple and four-choice reaction time; (Deary et al., 2011))
7. Alice Heim 2 test of general cognitive ability (Heim et al., 1974a,b), containing verbal, numerical, and pictorial reasoning items
8. Edinburgh Handedness Inventory (Oldfield, 1971). (Relevant to hand scan measurements)
9. Musical Learning study

### 4.2.3 CHOICE OF STIMULI FOR STUDY

The choice of musical stimuli for this study was one of the most crucial aspects of its design. These stimuli must be easy enough to be perceived and judged by many, but difficult enough that only the individuals with the highest ability can learn to reproduce them accurately. They must, furthermore, fit into a supposed "culture". Thus, great care was taken over the selection of the stimuli to be used. The following sections will outline initial investigations into artificial grammar learning, before describing the stimuli chosen and the basis for their choice.

#### 4.2.3.1 ARTIFICIAL GRAMMAR REVIEW

The Artificial Grammar Learning (AGL) paradigm is concerned with addressing what form of knowledge is acquired as a result of learning, and also whether this knowledge is implicit or explicit. The common paradigm employed is that of a finite-state artificial grammar. This is used to "specify a set of continuation relations among symbols, including beginning and end states, so that sequences of symbols can be constructed" (Pothos, 2007, 227).

The AGL paradigm divides the resulting sequences into those consistent with the finite state language, and those which are not (grammatical and ungrammatical, respectively). One of the first, and most influential, studies to use this paradigm is reported in Reber (1967). Reber created a finite state language by defining the letters P, S, T, V, X as the vocabulary, and a set of sentence construction rules as the grammar - see Figure 4.3. The maximum length for a "sentence" string was eight characters, while all fell between six and eight characters. Each participant was required to learn and reproduce a randomly-chosen set of 28 "sentences", presented as seven sets of four. There were five basic sentence types, here shown with loops placed within parentheses:

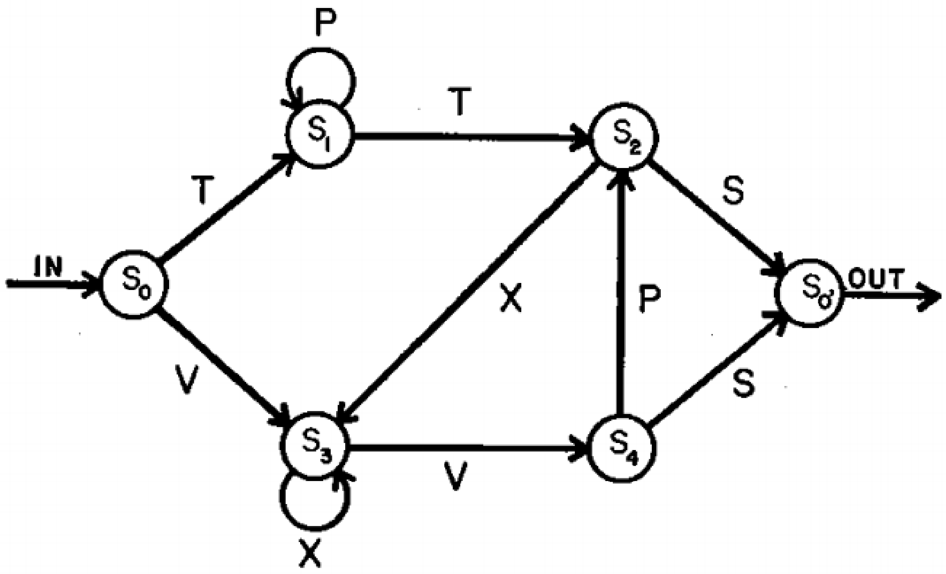


Figure 4.3: Original finite state grammar machine (Reber 1967)

a) T(P)TS

b) T(P)TX(X)(VPX(X))VS

c) T(P)TX(X)(VPX(X))VPS

d) V(X)(VPX(X))VS

e) V(X)(VPX(X))VPS

Reber's initial finding showed a consistent decline in error rates across the seven sets of sentences as learned by participants, which suggests that the participants were learning the grammatical relations inherent in the sentences. Reber (1967, 859) posits that "those in the experimental group were learning to be more efficient [than those in a control group, who had been exposed to a random language built from the same five letters], presumably due to the grammatical structure of their items". Reber also found that his participants had little to no verbalisable knowledge of the construction rules of the language. This indicates that the learning was implicit, rather than explicit.

Reber's experimental paradigm has been used to investigate a multitude of questions in language processing, sequence learning processes, and other fields. Previous experiments using music have used artificial grammars such as FSG (finite state grammar) and PSG (phrase structure grammar). Neuhaus et al. (2009), used both, translating the FSG form ABAB into interval changes of 3rd up/ 3rd down/ 6th up/ 6th down, and the PSG form A[AB]B into 3rd up/ 6th up/ 6th down/ 3rd down. Here, a "3rd up" refers to a leap up of two musical tones, while a "6th down" refers to a drop of six musical tones. However, other studies have used the classical Reberseque grammar machine to create grammars of musical timbres (Bigand et al., 1998), and contemporary dance moves (Opacic et al., 2009).

One problem with the application of generative grammar to music is the lack of structural musicology - whereas Chomsky's generative-transformational linguistics was built upon thirty years of structural linguistics, there has not been a corresponding musical background. As stated by Baroni et al. (1983, 183), "although it had used mature theoretical ideas, it had never cultivated the empirical clarity that this use [generative grammar] presupposed". Thus, musical grammar analysis is playing a catch-up game to linguistics. This is striking when one considers the vast impact of Schenker's musical analysis on musicology (Schenker, 1954, 1979), and the striking coincidences between Schenker's views and Chomsky's, particularly as Schenker was writing his most mature theories when Chomsky was still an infant.

Ultimately, I decided not to use an Artificial Grammar to create stimuli for this study due to reservations about whether a rule-based system would accurately create *cultural* pieces of music. I wanted to base my stimuli on a "miniature culture", from which I could be satisfied that participants might implicitly pick up on shared traits. I decided to use a system of music which has been analysed by ethnomusicologists, and which already constitutes a series of musical variations. The fact that these variations are all based on one theme, and are filtered through a real culture, constitutes a cultural group of musical practice, a "miniature culture". Potential examples of musical "miniature cultures" are the Javanese Gamelan cycle, and the rudiments of army drumming. However, North Indian Tabla was used, as variations on a theme were readily available and could easily be translated to the two tones chosen for this study. Additionally, it was expected that the majority of experimental participants would be unfamiliar with this music, as it is relatively confined to North India. Furthermore, the transposition of the complex tonal interplay of tabla music to just two tones was expected to remove a lot of the elements which might potentially make this music familiar to participants.

#### 4.2.3.2 TABLA

Tabla is a North-Indian drumming tradition about which relatively little is known in the West. This tradition consists of two hand-drums (tabla and baya), each with different timbres. Different tones and timbres can be made according to how and where each drum is hit. The word Tabla seems to derive from the Arabic term tabl, which means "drum". The variety of sounds which can be created are represented by mnemonic syllables, known as bols. Players of tabla intuit that these mnemonic syllables resemble the drum sounds (Patel and Iversen, 2003, 63). Bols are small units which can be combined in a multitude of ways to create new pieces of tabla music. This is analogous to the units of language. Readers are recommended to read Patel (2008, 34-36, 63) for a discussion of Tabla, and illustrative Tabla drum sounds.

Kayda are the formulae and rules upon which Tabla pieces are created. These are also important to the pedagogy of Tabla. They enable the creation of many variations on one theme. Tabla themes were chosen as fitting our purposes as they come from one identified musical culture, they are unlikely to be very familiar to the population under study (University of Edinburgh students), and, crucially, they all relate together through their underlying rules and structure. A beginner's TiReKiTe Kayda was chosen, and adapted for use in the study. The kayda from which the stimuli were adapted were taken from the website of the Indian Music Society of Minnesota. This source was chosen, as it presented class notes as used by contemporary Tabla musicians, rather than the possibly sterilised or wrongly-transcribed Kaydas recorded by ethnomusicologists. This provided us with one Kayda theme, and 13 variations, with a concluding *chakkardar tihaai*, or theme, which provided three further elaborations on the initial Kayda. This provided us with 16 possible related themes.

Whilst Tabla themes are richly nuanced musics with multiple sounds being made on each drum according to the tension and location of the hand impact, the aim here was simplification. Thus, just the rhythm was retained, of eight variations on a theme. These stimuli rhythms were synthesised using Garageband, a Mac application, using two woodblock sound effects. All rhythms were 6 or 6.5 seconds long, and set at 120 BPM (beats per minute). The "TiReKiTe" rhythms (roughly corresponding to four quavers) were always played on the lower of the two woodblocks, while the "Dha" and "Ti" (roughly corresponding to crotchets) alternated between the high and low block sounds. Eight rhythms were used in total, with an increasing complexity (as defined by rate of rhythmic activity, the measure of complexity used in experiment one, inspired by Conley (1981)). The westernised notation of these eight rhythms can be found in Figure 4.4. Our use of the term “rhythm” follows that of Fitch and Rosenfeld (2007) and Windsor and Desain (2000), as a temporal pattern of acoustic elements which occur with some repetitive structure.

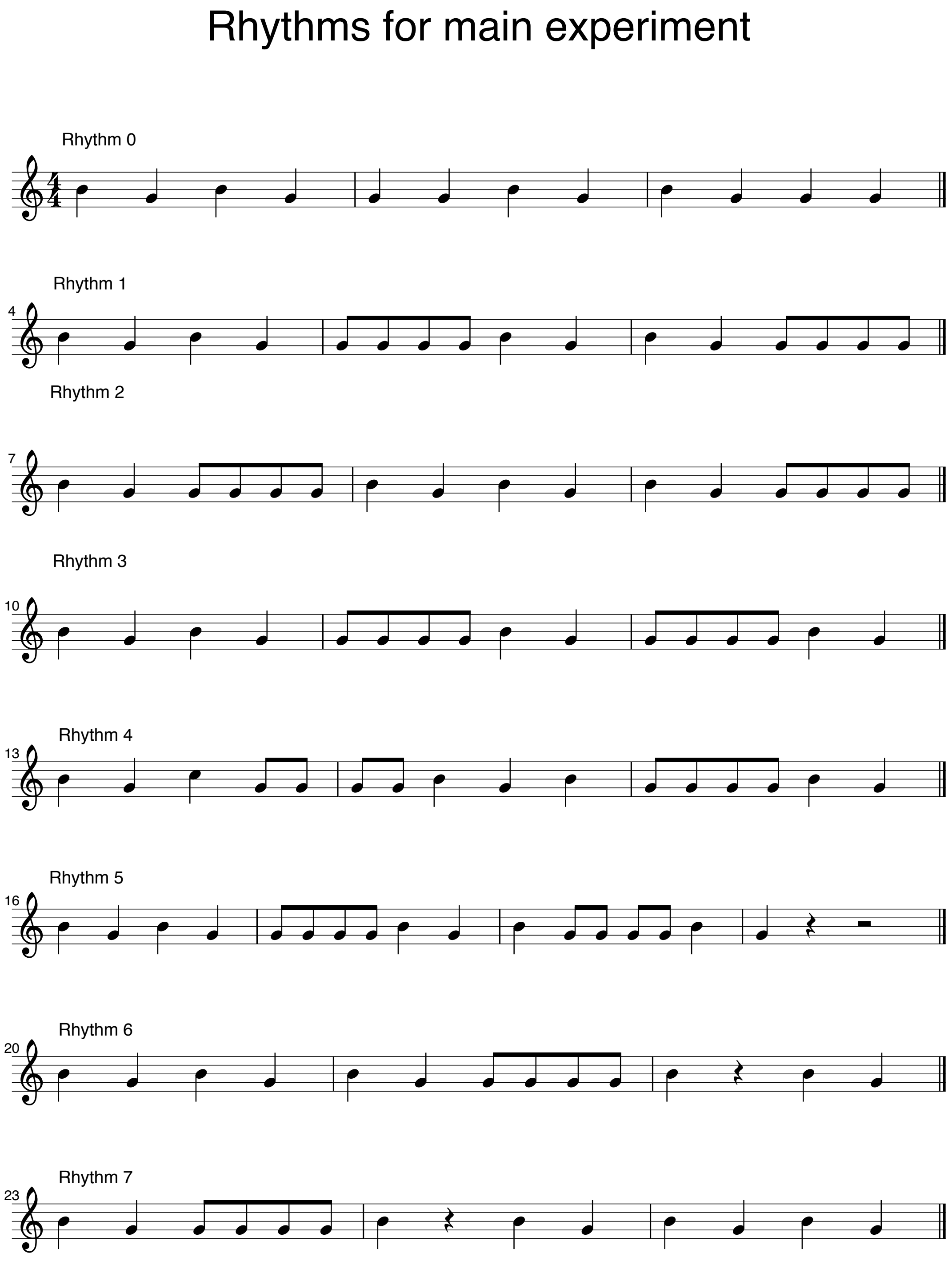



Figure 4.4: Rhythms used in study two, adapted from North Indian Tabla

### 4.2.4 MUSICAL LEARNING STUDY

In this stage of the study, participants were tested on their ability to learn a novel musical system - this constitutes the musical learning aspect of the study. This consisted of four phases. In phase one, they were asked to familiarise themselves with the musical instrument. This phase also allowed them to tap an isochronous rhythm without support from a metronome, providing additional time for the participants to become used to the novel instrument. The next three phases made up the main body of the study, and tested learning ability.

The second phase of the study tested each participant's ability to learn to imitate others. Participants listened to eight rhythms, each played four times. After each playing of each rhythm, they were asked to replicate it using a novel instrument. This instrument was an AKAI LPD8 MIDI laptop controller, a small machine which controls sound generation through the USB port of a laptop computer. The LPD8 has eight touch-pads, which were set to the two tones which were used in this study. This provided us with 32 replications for each participant - the four versions of each rhythm allowed us to observe any increase in ability across rhythms. In the third phase of the study, the participant was asked to recall the rhythms they had just heard, as many as they could remember. Finally, in phase four, they were asked to create rhythms which fit the style of the rhythms they heard during the study. This tested their ability to learn to fit into a cultural norm while making something creative.

These final three phases (the main body of the study) allowed us to observe each participant's learning ability on three key dimensions: ability to directly replicate the musical behaviour of others, ability to recall later this musical behaviour, and, thirdly, the ability to spontaneously create musical behaviours that fit the cultural norm. These three test phases address the three key elements seen above (section 2.1.1 on page 41) motor skill, cultural learning and creativity. Motor skill is tested by ability to learn to use the novel musical instrument, cultural learning is tested by the replication and recall rounds, and creativity in production by the creativity round. This experimental set up is not dissimiliar from that of Fitch and Rosenfeld (2007), who had participants tap in time along to rhythms, reproduce these rhythms given a steady pulse, and recognise these rhythms both immediately after presentation and after a 24 hour delay. However, the rhythms used by Fitch and Rosenfeld (2007) increased in complexity according to syncopation along a metric devised by Longuet-Higgins and Lee (1984), while the rhythms used in this study increased in complexity according to number of rhythmic elements, as seen in section 4.1.1 on page 93. Further, this study requires participants to create novel rhythms.

The study was run using a Processing script written by myself, on a MacBook running OSX Leopard. Participants listened to musical stimuli using Sennheiser headphones. This study was approved by the University of Edinburgh psychology department ethics forum. All participants signed consent forms to take part in the study, and for their data to be stored anonymously on a secure server. Full instructions given to participants can be found in the Appendix, on page 160.

#### 4.2.4.1 STUDY PLAN

1. FA test - hand, face, and body
2. IQ test (AH2)
3. Information processing speed test (Deary Liewald)
4. Start of musical study
    - Familiarisation with novel instrument
    - Isochronous rhythm familiarisation: metronome drops out partway through
    - Exact reproduction - replications of 8 rhythms
    - Recall of 8 rhythms
    - Creativity - 4 rhythms to be created
5. End of subject's participation - Post-test questionnaire

This study allowed us to draw connections between intelligence and fluctuating asymmetry, an index of developmental stability, in order to begin to examine the sexual vs social selection distinction. This is an explicit analogy to birdsong and the developmental stress hypothesis. So, we can investigate whether the proposed melodo-rhythmic precursor was an honest indicator of cognitive ability. This may address the question raised in chapter two, about the proposed relationship between cognitive ability and musical learning ability (see page 65). If the system underwent sexual selection, we can predict that FA would negatively correlate with intelligence and musical learning ability and that intelligence and musical learning ability would positively correlate. We would expect that more symmetrical participants (high developmental stability) would have higher IQ and quicker processing speed. This finding would support the hypothesis that the ability to learn to produce complex multi-modal, musical sequences could have signalled cognitive ability and developmental stability. However, if FA does not correlate with measures of cognitive and musical learning ability, then we might conclude that the LSP underwent social selection, in which physical abilities are irrelevant.

### 4.2.5 Modes of Analysis

As described above, the music learning study required participants to learn under three conditions, the first being replication, the second recall and the third creativity. These three different modes of learning musical sequences require three distinct methods of analysis, which will be discussed below.

#### 4.2.5.1 Replication Analysis - Edit Distance

The edit distance between two strings (sequences) is the minumum number of operations (substitutions, deletions, insertions) needed to transform one into the other. Essentially, it can give us a measurement of the difference between sequences. In this study, we analysed the difference between participants' replicated rhythms and target rhythms by calculating the edit distance, in other words how successful each participant had been in learning how to use the novel musical instrument to replicate musical pieces.

This measure allowed us to separate two important aspects of learning - note and timing. This analysis[2] gave us a Note Distance measure - the minimal number of notes to be inserted, deleted or substituted to transform the participant's version to the target, and a Time Distance measure - the minimal number of milliseconds inserted, deleted or substituted in order to transform the participant's version to the target. See Figures 4.5 and 4.6 for a graphical representation of these operations. These measures allow us to "flatten out" the timing and note analyses, separating each in order to analyse them separately. Of course, two identical strings, when analysed using this procedure, would return Note Distance and Time Distance values of 0 - as no note or timing differences exist.

#### 4.2.5.2 Recall analysis

In order to calculate how many rhythms had been remembered during the recall section of the musical learning test, the Edit Distance analysis was combined with a threshold of error. This involved using the edit distance measures outlined above (section 4.2.5.1), comparing each recalled rhythm to each template/target rhythm (from the original eight learned), and creating a matrix of recalled rhythms versus template rhythms. A threshold of error was created by hand-

---

[2] This Edit Distance measure was written by Professor Simon Kirby, and can be seen in the Appendix, on page162

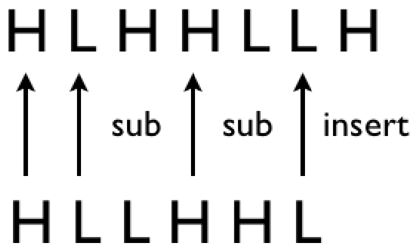


Figure 4.5: Representation of the edit distance operation - note values
The top line of note values (H = high, L = low) is the target, the bottom line of note values is the output created by the participant. In order to transform the output to the target, three operations are necessary, two substitutions and one insertion.

coding a subset of the data and identifying whether it matched a template rhythm. This involved listening to each rhythm and rating how well it matched each potential template rhythm. Raw error scores coincided remarkably well with intuitions, with lowest errors coinciding with intuitive matches. A threshold was created, of 0 to 1 note errors, and 0 to 2700 time errors (time errors were measured in milliseconds). In other words, if a participant recalled a piece to within 1 note error and 2.7 seconds of timing error, it was marked as a recalled rhythm. This resulted in a simple number between 0 and 8 for how many rhythms each participant recalled. Figure 4.7 shows the approach taken, using the matrix.

#### 4.2.5.3 Creativity Analysis

This thesis considers creativity to be "fitting but novel". In other words, a creative musical piece must be both cultural and interesting, new - it will take aspects of a musical culture, and add new elements, to make something which is neither slavishly cultural nor unlike anything which has gone before. As such, both elements - culture and novelty - must be addressed in the analysis. This discussion of both the novel and the cultural is relevant also to language, in which we find that interesting sentences balance the novel and the cultural, in other words are new but grammatical, creative within pre-determined cultural conventions. This idea of "fitting but novel" is not itself new. Guilford (1950, 453) states that "the creative person has *novel* ideas" and elucidates further that "creative work that is to be realistic or accepted must be done under some degree of evaluative restraint". Here, Guilford seems to be referring to both novelty and cultural fit. In 1953, Stein (1953, 311) defined the creative work as "a novel work that is accepted as tenable or useful or satisfying by a group in some point in time". He writes that "it

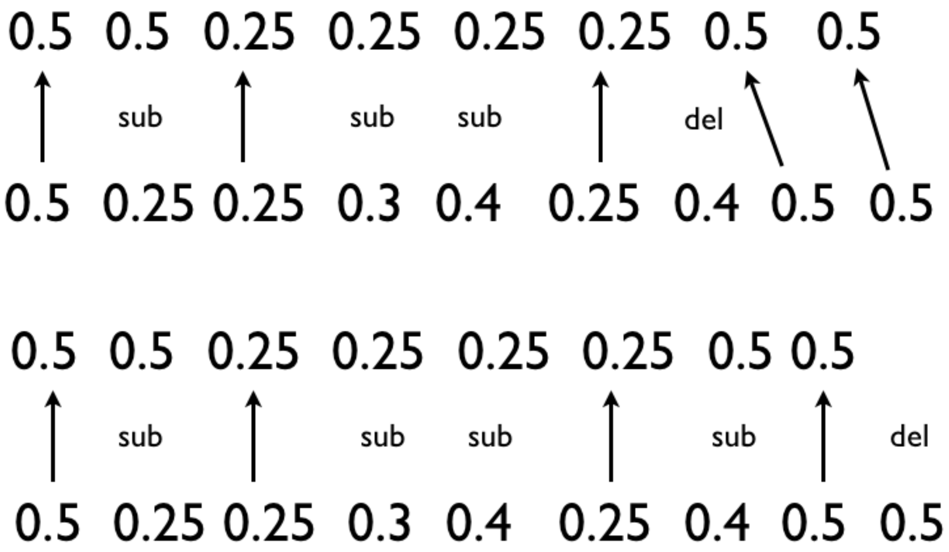
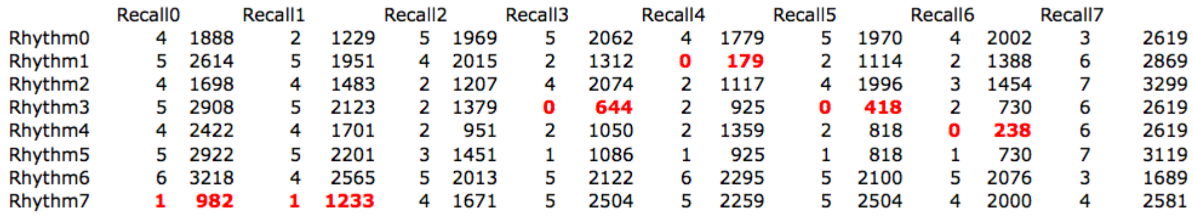


Figure 4.6: Representation of edit distance operation - time values
The top line of time values in each set (0.5 = half second, 0.25 = quarter second for example) is the target, the bottom line of time values is the output created by the participant. In order to transform the output to the target, two alternative edits are shown, with the top set needing least operations (three substitutions and one deletion), and so selected as the edit distance value (4).

arises from a reintegration of already existing materials or knowledge, but when it is completed it contains elements that are new".

Bruner (1962) also discusses creative products as those which produce "effective surprise" in the observer, in addition to a "shock of recognition". Thus, the product is novel but fitting - appropriate to a cultural system. This view of creativity was discussed by Mayer (1999) who summarised multiple definitions of creativity as "creativity involves the creation of an *original* and *useful* product". Thus, in our analysis of creativity, we might expect to see few examples of creative output that are slavishly cultural - the endeavour of creativity is to balance the novel and the relevant, in order to create a product which is neither slavishly cultural, nor entirely novel.

Creativity is clearly an ill-defined thing. I have suggested a definition above, as "fitting but novel". Under this description, objective measures of analysis seem difficult. As Amabile (1983b, 17) has written, "the definition and assessment of creativity have long been a subject of disagreement among psychologists". Amabile discussed various approaches to creativity. Some of these include creativity "tests" - personality inventories (Rimm, 1976; Rimm and Davis, 1976), biographical inventories (Schaefer, 1969; Taylor, 1963), behavioural tests (Guilford, 1967; Torrance, 1962), and environmental influences on test performance (Speller



| | Recall0 | | Recall1 | | Recall2 | | Recall3 | | Recall4 | | Recall5 | | Recall6 | | Recall7 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rhythm0 | 4 | 1888 | 2 | 1229 | 5 | 1969 | 5 | 2062 | 4 | 1779 | 5 | 1970 | 4 | 2002 | 3 | 2619 |
| Rhythm1 | 5 | 2614 | 5 | 1951 | 4 | 2015 | 2 | 1312 | **0** | **179** | 2 | 1114 | 2 | 1388 | 6 | 2869 |
| Rhythm2 | 4 | 1698 | 4 | 1483 | 2 | 1207 | 4 | 2074 | 2 | 1117 | 4 | 1996 | 3 | 1454 | 7 | 3299 |
| Rhythm3 | 5 | 2908 | 5 | 2123 | 2 | 1379 | **0** | **644** | 2 | 925 | **0** | **418** | 2 | 730 | 6 | 2619 |
| Rhythm4 | 4 | 2422 | 4 | 1701 | 2 | 951 | 2 | 1050 | 2 | 1359 | 2 | 818 | **0** | **238** | 6 | 2619 |
| Rhythm5 | 5 | 2922 | 5 | 2201 | 3 | 1451 | 1 | 1086 | 1 | 925 | 1 | 818 | 1 | 730 | 7 | 3119 |
| Rhythm6 | 6 | 3218 | 4 | 2565 | 5 | 2013 | 5 | 2122 | 6 | 2295 | 5 | 2100 | 5 | 2076 | 3 | 1689 |
| Rhythm7 | **1** | **982** | **1** | **1233** | 4 | 1671 | 5 | 2504 | 5 | 2259 | 5 | 2504 | 4 | 2000 | 4 | 2581 |

Figure 4.7: Example of threshold recall analysis
This matrix compares each recalled rhythm, as created by the participants, with each "target" rhythm. The numbers correspond to the note and time distance between each recalled rhythm and each target. Numbers in red represent recalled rhythms, as they have passed the threshold test of 1 note error and 2700 time errors

and Schumacher, 1975; Manske and Davis, 1968). These tests will be discarded, as they do not test the creativity of a *product*, but of *people*. As Amabile (1983b, 25) critiques, the validity of these tests is questionable, and this problem is further confounded when creativity measures are validated against one another. Furthermore, some of these tests may be influenced by verbal abilities - Torrance's test of Creative Thinking has been shown to be influenced by verbal fluency (Dixon, 1979; Hocevar, 1979). Amabile (1983b, 25-26) argues that "creativity researchers may have been too quick in attempting to objectify the assessment of creativity", suggesting that judgements of creativity "can ultimately only be subjective". She recommends that researchers "abandon the hope of finding objective ultimate criteria" (Amabile, 1983a, 359) for creativity measurement, instead adopting subjective criteria.

One striking *objective* study of music and creativity has been carried out by Simonton (1980a,b), who sought to investigate the relationship between a melody's fame and its originality. Simonton took 15,618 themes by 479 classical composers, using the first six notes of each theme to determine its originality. He paired each note with its successive neighbour, giving 5 two-note transitions for each six-note theme. Using computer analysis, each note transition was scored both on the rarity of the notes and the intervals between them. Here, rarity was defined as the negative of the percentage of times the note transition occurred (Simonton, 1980b, 975). Using this approach, Simonton studied melodic originality relative to both the entire repertoire (all themes), and the Zeitgeist at the time of composition. Multiple regression analyses showed that thematic fame is an inverse function of repertoire melodic originality, Zeitgist melodic originality, and a positive function of creative productivity. Repertoire melodic originality is a positive function of time and an inverted function of composer's age, and Zeitgeist melodic originality is a positive linear function of composer's age. Whilst this approach is seemingly entirely objective (indeed, Simonton (1980b, 972) states that "music is the most mathematical of the arts...accordingly a theme's melodic originality can be assessed in an entirely objective and reliable manner"), due to time constraints and reservations about the measurement of "Zeitgeist", this measure was not used. It may prove fruitful for future work.

Amabile (1983a, 360) states that something will be judged creative "to the extent that a) it is both a novel and appropriate, useful, correct, or valuable response to the task at hand, and b) the task is heuristic rather than algorithmic". Thus, Amabile defines creative products as novel but fitting, stating that novelty and appropriateness are two hallmarks of creativity. Further, she states that heuristics must be used to solve a creative task - no readily identifiable path to solution should be in place.

When discussing precisely how to judge the creativity of a product, Amabile writes that

> "A product or response is creative to the extent that appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated. Thus, creativity can be regarded as the quality of products judged to be creative by appropriate observers" Amabile (1983b, 31).

This is a consensual definition, based on the creative *product*, and so, is useful for our analysis of creativity. As Amabile writes, identifying a thought process or subprocess as creative must depend upon there being a creative *product*. So, "the definition that is most likely to be useful for empirical research is one grounded in an examination of products" (Amabile, 1983b, 31). Amabile's approach to creativity rating here, asks the "appropriate observers" to rate products of a creative enterprise "using your own subjective definition of creativity" (Amabile, 1983b, 42, 52). This is the approach taken in the following section.

Subjective ratings are, by virtue of their subjectivity, difficult to quantify. As Amabile (1983b, 29) writes, "what, exactly, do judges mean when they call something 'creative'? What features of products predict their responses? What phenomenological response states lead them to apply that label?". Again, it is outside of the scope of this thesis to address these questions definitively, however, some attempt was made, in informal rater debriefings, to address raters' approach to the question of creativity.

**Creativity Rating** Twenty raters (the "appropriate observers" as defined by Amabile (1983b)) were recruited to rate both the "cultural fit" and "creativity" of the creative pieces participants had made. Cultural fit might be seen as analogous to cultural adherence or conformity. Each rater was trained by listening to the eight original "target" rhythms (the rhythms each participant had been asked to replicate and recall), then asked to rate each rhythm made by participants. Only one "creative" rhythm was taken from each participant, as representative of their creativity. Each participant had been given the opportunity to create four "creative" rhythms in the third section of the musical learning test. The second one was chosen to be rated, as the first may have behaved as a practice round of sorts. Time and financial constraints meant that only one creative rhythm could be rated. The rating was conducted in ten short blocks. Five of these blocks concerned how culturally appropriate a given rhythm was - raters were told that their job was to rate how successful participants had been at making something that sounded like the eight original rhythms. The five other blocks concerned novelty - raters were told to rate each rhythm according to how creative they thought each one was. They were informally told to use whatever their own internal method of creativity is, and, if they questioned further, it was suggested that perhaps this might mean that the rhythms were "interesting, new, or exciting". Each block was identical, with the rater hearing the eight original rhythms, then rating 20 (in the final blocks, 17) rhythms according to either cultural fit or creativity. Rhythms within the blocks were randomised, as was the order of presentation of the blocks. However, the five culture rating blocks were conducted one after another, as were the creativity blocks. Half of the raters rated creativity first, half cultural fit. The order of this study is shown in Figure 4.8.

In informal debriefs, some raters commented that they had rated the creativity as whatever deviated from the original eight rhythms - whatever was "new". Furthermore, some raters commented that rating creativity was almost the opposite to rating cultural fit. This seems to fit our concept of novelty in creativity.

### 4.2.6 RESULTS

This study aimed to investigate the relationships between musical learning ability, cognitive ability and developmental stability. Statistical analyses will be presented which support the LSP hypothesis discussed in chapter two. Data consisted of symmetry measurements of the face,

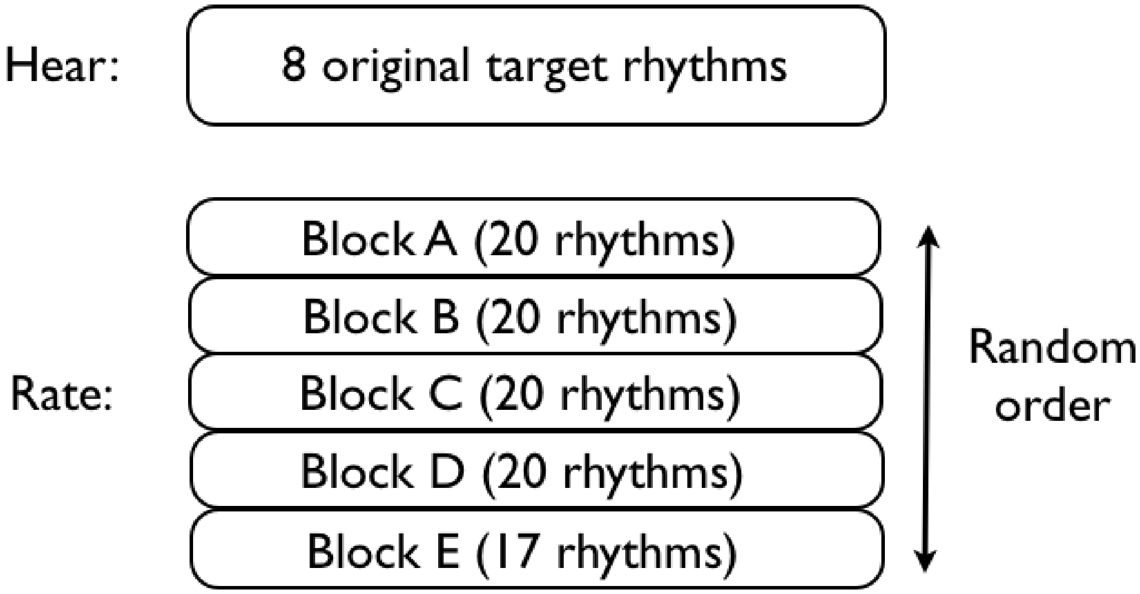


Figure 4.8: Order of presentation of creativity rating
Half of participants rated each rhythm on its creativity first, half on its cultural adherence

hands, and body, intellience test results (verbal, numerical and perceptual) and processing speed (simple or choice reaction time), alongside analysis of participants' ability to replicate (edit distance measures), recall (threshold test based on edit distance), and make creative pieces (as rated by naive raters).

#### 4.2.6.1 Replication Data

All rhythms were compared to their respective templates using an edit distance measure. This edit distance measure calculated the smallest number of substitutions, deletions and insertions needed to transform each participant's rhythm into the template. This resulted in two measures for each replication - a Note Distance value, and a Time Distance value. These values were visually inspected, and any values significantly above or below the average (note distance mean = 2.18, s.d. = 2.4, time distance mean = 1435.52, s.d. = 1193.53) error checked. Error checking consisted of listening to the audio samples of the relevant rhythm, and noting any false starts. Following this analysis, 15 false starts were removed, only from the very first rhythm replicated (rhythm 0 as seen in Figure 4.4). This was implemented as this is such a novel task that participants made errors and restarted most often in rhythm 0. Additionally, some participants missed some replications, clicking through the program without playing each rhythm four times. In these cases (40 in total), the average error for the remaining versions of this rhythm was taken as the error for the missing one(s). Further to this, both measures were Z-standardised, and values above 3 or below -3 were deleted (3 standard deviations from the mean).

The data was checked with a Kolmogorov-Smirnov test, which notes whether the distribution of data deviates significantly from normal. Data that were found to be not normally distributed included averaged note errors, averaged timing errors, number of rhythms recalled, averaged creativity ratings, years of musical experience, mean CRT (choice reaction time responses), mean SRT (simple reaction time responses), FA of the face, and body.[3] Visual inspection of QQ plots confirmed that these data were non-normally distributed. As the normality of distribution assumption was not met here, non-parametric tests will be presented throughout this analysis.

It was predicted that there would be a positive correlation between error as measured by note distance and by time distance, and so the correlation between note distance and time distance measures was checked by using the Spearman's correlation coefficent. This coefficient will be used throughout this chapter. There was a significant correlation between error as measured by averaged note distance and by averaged time distance ($r_s$(102)=.93, p<.005). So, participants who made more timing errors also made more note errors.

In order to test the reliability of our two measures of error, the Cronbach's alpha coefficient was conducted on both. This measures the consistency of error rates across stimuli. The data were averaged across versions of each rhythm giving us an average error for each rhythm for each participant, according to the two measures of error, Note Distance and Time Distance. Analysis revealed that both measures had a high reliability, with a Cronbach's alpha (standardised) of 0.91 for Note Distance, and a Cronbach's alpha (standardised) of 0.89 for Time Distance. Cronbach's alpha should be at least 0.60, if not 0.8, and so both measures are highly reliable.

As each participant played each rhythm four times, we can observe any improvement in per-

---

[3] p values here are highly significant and too exhaustive for in-text discussion; Data found to be non-normally distributed included averaged note errors (D(72)=.154, p=.000227), averaged timing errors (D(72)=.155, p=.000199), number of rhythms recalled (D(72)=.215, p=.000), averaged creativity ratings (D(72)=.115, p=.019), years of musical experience (D(72)=.146, p=.000013), mean CRT (choice reaction time responses) (D(72)=.109, p=.034), mean SRT (simple reaction time responses) (D(72)=.130, p=.004), FA of the face (D(72)=.109, p=.033), and body (D(72)=.126, p=.006

formance across subsequent productions. In order to do this, a Friedman's test was conducted. The two measures of analysis were analysed separately. Data were averaged across rhythms, to give an average score for each participant, by the first, second, third and fourth versions of each rhythm.

In order to investigate the role of rhythm number (which may be related to complexity - see section 4.2.6.2 below) on the error rates, data were averaged to give an overall error score for each version of all rhythms. Mean note error fell between version 1, version 2, version 3, and version 4. The Friedman's test revealed a significant effect of version on mean note errors ($\chi^2$(3)=141.30, p<.001). Please see Table 4.4 below.

Mean time error fell between version 1, version 2, version 3, and version 4. The Friedman's test again revealed a significant effect of version on time distance scores ($\chi^2$(3)=87.66, p<.001), which can be seen in Table 4.5 below.

| Version | Mean | Std. Dev. |
|---|---|---|
| V1 | 3.12 | 1.75 |
| V2 | 2.38 | 1.78 |
| V3 | 1.87 | 1.61 |
| V4 | 1.74 | 1.73 |

Table 4.4: Mean note distance errors according to version of rhythm
There was a significant effect of version on note distance (p<.001)

| Version | Mean | Std. Dev. |
|---|---|---|
| V1 | 1900.32 | 937.49 |
| V2 | 1545.90 | 905.09 |
| V3 | 1383.46 | 889.95 |
| V4 | 1319.78 | 974.43 |

Table 4.5: Mean time distance errors according to version of rhythm
There was a significant effect of version on time distance (p<.001)

The Friedman's test revealed significant effects of version on error scores - in other words, error varied according to the version of each rhythm played, with less error for later versions. Unsurprisingly, the more exposure participants had to each rhythm, the better they were at replicating it, and the lower their note and time errors.

#### 4.2.6.2 INCREASING COMPLEXITY ACROSS RHYTHMS

The rhythms used in this study were adapted from Indian Tabla patterns - a TiReKiTe pattern with variations built upon a central theme. This building up of variation was taken as an index of increasing complexity. Furthermore, these rhythms fit Conley's (1981) discussion of rate of rhythmic activity as a measure of complexity (as discussed earlier in this chapter). It was expected that participants would find each rhythm increasingly more difficult to replicate, and that this would be reflected by higher note and timing errors for later rhythms. However, whilst there was a general move from less to more errors according to increasing rhythm number, there was an interesting "see-saw" pattern from rhythm to rhythm. This is clearly seen by looking at the median error values of each rhythm, shown in Table 4.6. .

| Rhythm | Note errors | Time errors |
|---|---|---|
| 0 | 2.50 | 1533.50 |
| 1 | 1.00 | 896.25 |
| 2 | 2.00 | 1255.00 |
| 3 | 1.88 | 1374.50 |
| 4 | 2.00 | 1282.25 |
| 5 | 1.75 | 1159.00 |
| 6 | 2.00 | 1519.75 |
| 7 | 1.25 | 1219.50 |

Table 4.6: Median values of note and time errors made on each rhythm.
Rhythm numbers used here apply to the rhythms shown in Figure 4.4

**Discussion** This analysis has shown that the two measurements of error, note distance and time distance, correlate significantly, and are reliable and consistent measures of musical learning ability. Furthermore, participants vary in their ability to learn the rhythmic pieces (as measured by note and timing errors). The number of rhythm had a significant effect on the error scores, and there is a significant difference between different versions of each rhythm (as we have eight rhythms each played four times by each participant). If we follow the Conley (1981) analysis of complexity as rate of rhythmic activity, we may view increased rhythm number as indicative of increased complexity. So far, the music learning study seems to be an appropriate method of measuring people's individual differences in performance.

#### 4.2.6.3 RECALL DATA

As seen in section 4.2.5.2, number of rhythms recalled was evaluated using a threshold analysis. The mean recall number was 1.47, with a standard deviation of 1.26, and the mode was 1, within a range of 0 to 5. Recall number significantly negatively correlated with Note error averages ($r_s$(97)= -.60, p<.001) and Time error averages ($r_s$(98)= -.64, p<.001) from the replication section, indicating that participants who made less errors when copying rhythms also recalled more of them later.

**Creativity/ Cultural Adherence Data** The third part of the musical learning study required participants to make something creative. As seen in section 4.2.5.3, creativity here is seen as "fitting but novel". Thus, participants had to create something which was cultural and fitting, while still being interesting and novel. Raters were used to provide subjective analysis of both creativity and cultural adherence. Amabile (1983b) states that raters can be reliable and consistent when asked to rate according to their own subjective and personal idea of creativity. Participants (N = 20, M = 8) were recruited from the psychology subject pool and careers website at the University of Edinburgh, and were compensated with course credit or £6, as the study took on average under an hour.

Raters reliably provided creativity (Cronbach's alpha = .85) and cultural adherence (Cronbach's alpha = .93) measures for each participant, on the basis of one "creative" rhythm. These were significantly correlated ($r_s$(95)=-.63, p<.001). However, rated creativity did not correlate with any of the other measures of musical learning - Note Distance ($r_s$(95)=-.04, p=.72), Time Distance ($r_s$(95)=-.07, p=.49), or Recall Number($r_s$(95)=-.03, p=.80). Cultural adherence, on the other hand, did correlate with Note Distance ($r_s$(95)=-.22, p=.03), Time Distance ($r_s$(95)=-.21, p=.04), and Recall Number ($r_s$(95)=.21, p=.04). This indicates that, unsurprisingly, when raters are rating the cultural adherence of a musical piece, they look to the producer's ability to copy and recall pieces well. Creativity is, perhaps, not concerned with these aspects, but more with an *absence* of cultural adherence.

There was an interesting association between cultural adherence and creativity, as seen in Figure 4.9. Here, we can see that, according to ratings, one cannot be both creative and slavishly cultural - there is an "arms race" between cultural adherence and creativity, which mediates the musical output made by participants. However, rhythms that were rated low on the cultural

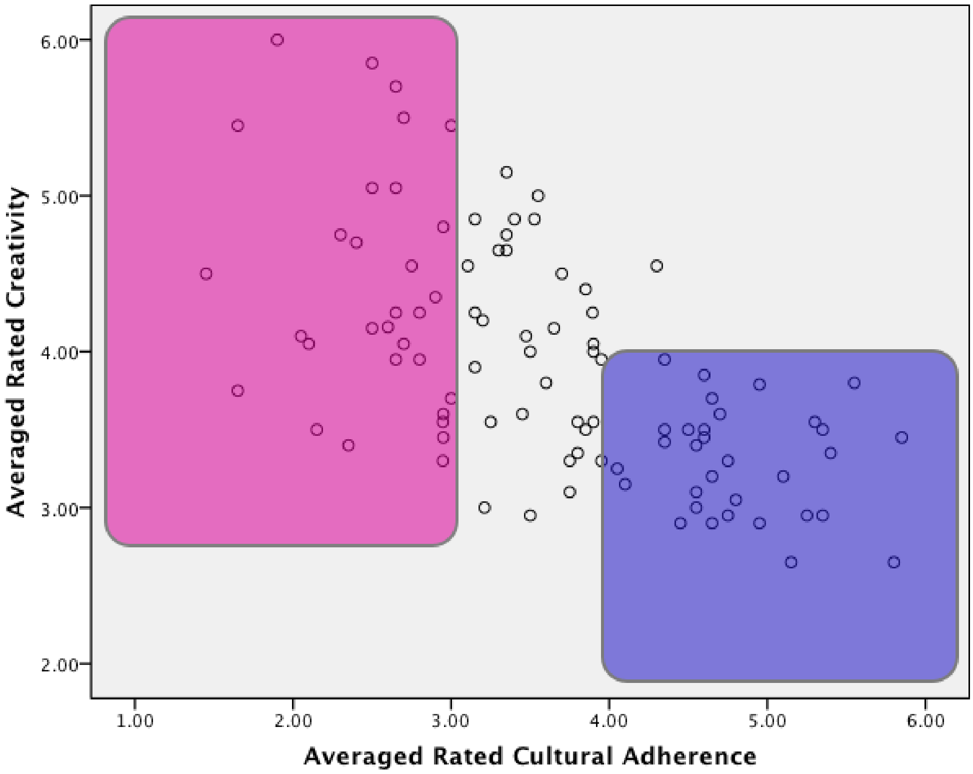


Figure 4.9: Scatter plot of creativity and cultural adherence
The blue section shows that one cannot be both highly cultural and creative, while the pink section shows the wide range of ratings of creativity possible when a piece is rated low on cultural adherance

adherence scale exhibited a wide variety of ratings in creativity - there was a larger scale of creativity rated for non-cultural pieces. This raises questions about the role of cultural adherence and creativity in a learned musical task, and, by extension, in language also.

#### 4.2.6.4 Tests of cognitive ability

In order to test general intelligence, participants were subjected to a battery of tests, including the Alice Heim 2 test of verbal, numerical, and perceptual intelligence. The test answers of each participant were marked according to the marking guide, by myself and experimenters from the psychology department of the University of Edinburgh. Significant correlations exist between all aspects of this test. Whilst many of the experimental participants were non-native English speakers, the Alice Heim 2 test was conducted in English. In order to control for any first language effects, first language was partialled out of the data using a linear regression and saving the standardized residual. This gave us a AH2 Verbal Residual score, which will be used in place of the original verbal data in this analysis.

Furthermore, participants also underwent the Deary-Liewald Reaction Time test, in order to test their information processing speed. This test consisted of two aspects - simple reaction time, which required participants to respond as quickly as possible to a cross (X) which appeared on the screen. They responded by pressing one button. In the choice reaction time test, however, participants were told that a cross might appear in one of four areas on the screen - left-most, middle left, middle right, right-most. Four buttons on the keyboard corresponded to the four possible areas. They must hit the corresponding button when the stimulus appeared on the screen. The time taken to respond to stimuli in each of the two tests informed our analysis of information processing speed. These two measures of processing speed (simple and choice) were significantly correlated ($r_s$(98)=.44, p<.001).

**Associations between cognitive ability tests** In order to investigate associations between these two measures of cognitive ability, correlations were calculated. No correlations were found between mean SRT (Simple reaction time) and verbal ($r_s$(99)=-.09, p=.39), numerical ($r_s$(99)=-.04, p=.68), or perceptual ($r_s$(99)=-.17, p=.09) ability. Choice Reaction Time (CRT) means correlated significantly with perceptual ($r_s$(99)=-.23, p=.02), but not verbal ($r_s$(99)=-.005, p=.96) or numerical ($r_s$(99)=-.16, p=.12) ability.

In order to identify *general* intelligence and processing speed from the data, these underlying variables were extracted using Principal Components Factor Analysis. This analysis is standard in the field of evolutionary psychology (Jensen, 1998; Penke et al., 2012). This provided us with two variables, identified as g speed (accounting for 49.9% of variance in the data) and g intelligence (accounting for 65.91% of variance in the data). g speed and g intelligence are significantly correlated ($r_s$(92)=-.24, p=.02). These measures will be used in subsequent analysis, unless more detailed analysis using aspects of the Intelligence and Processing Speed tests is warranted.

#### 4.2.6.5 FLUCTUATING ASYMMETRY

As detailed in section 4.2.2, FA data included hand scans (one scan of each hand, measured independently by two experimenters to establish reliability), facial symmetry (two versions) and body symmetry (three versions of a full body scan). In order to calculate the FA of the hands, measurements of the length of each digit were made, and compared bilaterally. This was performed on both versions of the scans in order to limit error.

Facial symmetry was calculated, using Morphanalyser software, by measuring the distances bilaterally from pre-set points on the face (23 facial landmarks). These points were positioned on the inside and outside corners of the eyes, cheekbones (widest horizontal part of the face below eyes), widest point of the nose (at sides of nostrils), corners of the mouth, the jaw (widest horizontal part of cheeks at the mouth), and the base of the chin. The calculation of 80 sets of distances gave us 160 total data points for each participant's face - these were further averaged across the two versions of each face scan. Measurements were made by pixel distance.

The body's symmetry was calculated,using the $TC^2$ NX-16 3D body scanner software, by comparing bilaterally multiple traits - thigh length, thigh, knee, under-knee, under-knee height, calf, mid length girth, biceps, elbow, forearm, wrist, shoulder slope, shoulder to wrist, bust to waist, shoulder to waist (front and back), back waist to mid shoulder, mid shoulder to bust, back-waist to shoulder, shoulder to front-waist, back-waist to side-neck. These calculations allowed us to tease apart different possible roles for body parts in affecting symmetry - we looked separately at body, hands, body and hands averaged, upper body, lower body, and body and upper body without hands.

Control variables such as handedness, ethnicity and the body scanner average deviation (average deviation of day-to-day calibration accuracy) had no significant effects on FA scores. Outliers more than four standard deviations from the mean were removed as were data points of parts of the body which had undergone trauma. This affected only 0.8% of the original data points.

The FA measurements were further corrected for Directional Asymmetry (DA), as there was found to be widespread DA in face and body. This was corrected for by Principal Components Analysis, as suggested by Graham et al. (1998) and Simmons et al. (2004). Directional asymmetry indicates a systematic bias towards greater trait expression on one side (Penke et al., 2009, 431), and arises due to "non-random genetic and/or developmental processes, and, as such, may not be revealing of underlying developmental instability" (Simmons et al., 2004, 864). This principle components analysis was conducted on the covariance matrix between measures of left and right sides, for each DA trait. First Principal Components Analyses represent the sums of the variation in trait sizes, whilst the second principle components analyses represent estimates of individual asymmetry (Graham et al., 1998).

#### 4.2.6.6 Association between key variables

The LSP outlined in this thesis predicts that musical learning ability should be correlated with cognitive ability. Further, this experiment aims to investigate the role of physical quality in the selection of this system - if sexual selection were key to the system, then we might expect to find correlations between FA and cognitive and musical learning abilities. However, if the LSP underwent social selection, as is hypothesised in this thesis, then no correlations between FA and other variables should be expected. This may shed light upon the role of a musical, learned cultural system as a signal of fitness. In order to test this, correlations were calculated between our three measures of musical learning (Replication, Recall, Creativity) two measures of cognitive ability (g speed and g intelligence) and Fluctuating Asymmetry. In post-test questionnaires, participants were asked about their musical experience.

This study was designed in order not to give trained musicians an "edge". However, correlations were found between years of musical experience (as reported by participants) and note ($r_s(97)=-.34$, $p<.001$) and time ($r_s(98)=-.41$, $p<.001$) errors, recall number ($r_s(100)=.34$, $p<.001$), g speed ($r_s(95)=-.22$, $p=.032$), and g intelligence ($r_s(97)=.38$, $p<.001$). Regressions also supported this finding, with note and time errors, recall number, g speed and g intelligence all predicted by years of musical experience. In order to control for musical experience going into the study, and to account for musical *learning* ability,the variable which noted years of musical experience was partialled out of analyses.

As much of the data were non-normally distributed, a standard (parametric) partial correlation was inappropriate. The Kendall's partial rank correlation coefficient was calculated on the data, allowing the examination of the relationship between variables, partialling out (controlling for) years of musical experience. Kendall's partial rank was calculated by hand, rather than using a statistical package, using the formula below (taken from Cramer (1997, 271). This formula for computing the partial rank correlation ($\text{tau}_{12.3}$) between two variables (1 and 2) partialling out a third (3) variable is the rank correlation between the first two variables ($\text{tau}_{12}$) minus the product of the rank correlation between variable 1 and variable 3 ($\text{tau}_{13}$) and the rank correlation between variable 2 and variable 3 ($\text{tau}_{23}$), divided by the square root of the product of 1 minus

$$tau_{12.3} = \frac{(tau_{12})-(tau_{13}\times tau_{23})}{\sqrt{(1-tau_{13}^2)\times(1-tau_{23}^2)}}$$

the squared rank correlation between variables 1 and 3 ($tau_{13}^2$) times 1 minus the squared rank correlation between variables 2 and 3 ($tau_{23}^2$).

Table 4.7 shows correlations between musical learning measures and measures of cognitive ability, controlling for years of musical experience. Correlations were found between g intelligence and note and time errors and number of recalled rhythms, and between g speed and time errors, and number of recalled rhythms. Creativity and cultural adherence did not correlate with either measure of cognitive ability. Based on these data, we can conclude at this stage that some measures of musical learning (note and time errors, and number of recalled rhythms) correlated with our two measures of cognitive fitness (g intelligence and g speed). This lends support to the hypothesised LSP.

| | Note Errors | Time Errors | Recalled Rhythms | Creativity | Culture |
|---|---|---|---|---|---|
| g intelligence | -.30* | -.29* | .28* | -.06 | .09 |
| g speed | .08 | .13* | -.21* | .03 | -.04 |

Table 4.7: Correlations between musical learning, g intelligence and g speed, controlling for years of musical experience
Note: *indicates that correlation is significant at 0.05

**Recall and Replication association** As seen above (section 4.2.5.2), our measure of recall ability correlated with our measures of replication ability, Note error ($r_s(97)=-.60$, $p<.001$), and Time error ($r_s(98)=-.64$, $p<.001$). This correlation, and the related functions of both tests, indicate that aspects of replication ability may affect recall ability. This hypothesis was, however, not supported when note and time errors (replication ability measures) were partialled out of correlations between number of recalled rhythms and g speed and g intelligence. This was conducted using a Kendall's partial rank correlation coefficient, as seen above. So, although numbers of rhythms recalled correlated with replication ability, this did not have any effect on the significant correlations between our measures of cognitive ability and number of rhythms recalled. See Table 4.8 for the partial correlation values.

| | g speed | g intelligence |
|---|---|---|
| Recall number (partialling out note errors) | -.21* | .21* |
| Recall number (partialling out time errors) | -.18* | .20* |

Table 4.8: Correlations between number of rhythms recalled (partialling out both note and time errors), g intelligence and g speed
Note: * indicates that correlation is significant at 0.05

**Fluctuating Asymmetry & other variables** In order to investigate the interplay between cognitive ability, musical ability and developmental stability, correlations were calculated between the variables (g speed, g intelligence, and FA). Previous studies have suggested significant associations between IQ and FA (Furlow et al., 1997; Bates, 2007; Prokosch et al., 2005). Due to the inter-correlated nature and multiple comparisons when correlating this data, the risk of type 1 error, i.e. of artificially inflating the likelihood of a significant result, was increased. As a result, Bonferonni's correction was used here to correct the alpha value to 0.00625. With this correction in place, no correlations existed between the measures of FA and the measures of musical learning, g intelligence or g speed.

One reason for the lack of findings here may be the group of participants under study. This study was performed on 104 undergraduate students from the University of Edinburgh. It may be the case that these students have not undergone enough developmental instability for such a perturbation to be expressed in their symmetry. These are largely middle class students in a first world country, and as such are highly unlikely to have suffered the kinds of developmental instability suffered, for example, by a child living in a slum in India. Further, these students are highly unlikely to have suffered the kind of instability likely undergone by humans throughout evolutionary history, living as they do in a world which modern man has suited to his needs. These are WEIRD (Western, Educated, Industrialised, Rich, Democratic) participants (Henrich et al., 2010), and as such may not be fully representative of the species as a whole. Indeed, as Henrich et al. (2010) identify, WEIRD participants are highly unusual relative to other cultural groups. However, this is a criticism that could be levelled at the majority of psychological and behavioural science studies, which use students recruited from US or UK universities.

Alternatively, this lack of findings may in fact lend support to the social selection hypothesis. As briefly mentioned above, as social selection does not rely on genetic quality (based as it is on reproductive success as a result of *social* interactions), then measurements of physical quality should not be relevant to choices made on a non-sexual, social basis. As predicted above, if social selection is key to the evolution of the LSP, then we should expect to find no correlations between music and cognitive abilities and FA. This prediction was upheld in this study.

| | Note Errors | Time Errors | Recalled Rhythms | Creativity | Culture | g speed | g intelligence |
|---|---|---|---|---|---|---|---|
| FA face & body | -.03 | -.01 | -.05 | -.07 | .16 | .03 | .10 |
| FA face | .05 | .07 | -.14 | -.02 | .10 | .10 | -.01 |
| FA body | -.10 | -.14 | .05 | -.004 | .09 | -.07 | .12 |
| FA lower body | -.05 | -.02 | .06 | .14 | -.07 | .02 | .06 |
| FA upper body | -.09 | -.08 | .02 | -.17 | .23 | -.12 | .15 |
| FA hands | .02 | .08 | -.04 | -.08 | .09 | -.03 | .08 |
| FA lower body without hands | -.10 | -.16 | .04 | -.20 | .23 | -.08 | .12 |
| FA upper body without hands | -.11 | -.15 | .10 | -.04 | .11 | -.06 | .13 |

Table 4.9: Table showing a lack of correlations between Fluctuating Asymmetry and musical learning, g intelligence and g speed

## 4.3 Discussion

This chapter has outlined two empirical tests of the predictions of the hypothesis proposed in chapter two. The first addressed the role of complexity in the hypothesised LSP. It was found that complex pieces were preferred by naive raters, and these raters found these 'musicians' more attractive and intelligent. Complexity is suggested as important to the LSP, as it provided a way of signalling learning ability.

The second study is proposed as the beginnings of a novel empirical paradigm, involving aspects adopted from musicology, psychology, and biology. If language evolved from a musical signalling system which acted as an honest signal of cognitive ability, we should find correlations between the variables of interest - cognitive ability and musical learning ability. This finding would support the theory that the LSP acted as an honest signal of cognitive ability, through the broadcast of complex musical signals. The study outlined above, conducted on 102 participants, found correlations between musical learning (on two of the three measurements -

replication and recall ability) and cognitive ability (both general intelligence and information processing speed). Creativity did not correlate with cognitive ability, which raises questions about the role of creativity and its effect on musical output. This may provide evidence against the Miller (2000) view that music evolved as a sexual selection mechanism. As found here, creativity does not correlate with cognitive ability nor fluctuating asymmetry (an inverse indicator of developmental stability). This argues that creativity in music may not have been sexually selected. It may also be the case that the measure of creativity used here is not comparable to the more robust, more commonly used measures such as FA and g intelligence, g speed. Correlations were found between cultural adherence and ability to replicate and later recall pieces. This indicates that, unsurprisingly, when raters are rating the cultural adherence of a musical piece, they look to the producer's ability to copy and recall pieces well. The fact that no correlations with FA were found supports the social selection hypothesis, that the LSP evolved through social selection.

The main finding of this study was that cognitive ability was shown to correlate significantly with musical learning ability, supporting the hypothesis that learned musical signals could have acted as a signal of cognitive ability. This provides clear evidence against the criticism discussed in chapter two (page 65). That section outlined a potential problem for the hypothesis that there was a necessary relationship between cognitive abilities and musical learning ability, that of musical savantism. My experimental finding of correlations between measures of cognitive ability and musical learning ability supports the hypothesised relationship, providing robust evidence against the savantism argument. Furthermore, as will be investigated more in chapter five, this hypothesis proposes that the learned musical sequential signalling system could have undergone social selection driven by collaborative partner choice, as perceivers of musical signals could use these as a cue to the fitness of a signaller as a *collaborative* partner. Cognitive ability is sufficient here if we consider that, in undergoing some collaborative task such as hunting, a useful collaborator will be able to plan ahead, will be innovative and thoughtful. The following chapter will look in more detail at the mechanism proposed to have driven the selection of the LSP, namely collaborative partner choice.

# CHAPTER 5

# SELECTION

## 5.1 STUDY THREE - MATE CHOICE AND COLLABORATIVE PARTNER CHOICE

Many previous theories (Darwin, 1871a; Miller, 2001) have theorised that sexual selection was the main or only selection mechanism which acted on the proposed protolinguistic systems. However, as hypothesised in chapter two, a social, collaborative mechanism may have operated on the proposed musical precursor. As the musical precursor behaved as an honest signal of cognitive ability, perceivers could use this signal as a cue to the abilities of each potential collaborator. Thus, both the signaller and perceiver would have benefitted from a collaborative enterprise - for example, hunting. The signaller would have won increased resources, which means they would live longer, enabling more fecundity, and allowing them to lavish more resources on their offspring. Thus, the increased cognitive ability which allowed this signaller to learn and reproduce complex sequences of musical sounds would have persisted in subsequent generations.

One reason to reject the sexual selection hypothesis is the findings of study two - particularly section 4.2.6.6 on page 134. FA and musical learning ability did not correlate in the way we might expect if musical ability was a signal of *genetic* quality. There were no robust correlations found between note and time errors, creativity, or cultural adherence and FA measurements. Miller (2000) states that music evolved as a result of sexual selection. However, this is contradicted by my findings. If, after all, this musical signalling system *were* undergoing sexual selection, we would expect to see interactions between musical ability and relevant aspects of genetic quality, expressed by fluctuating asymmetry, for example. This is not the case, and so, for the study of this proposed LSP certainly, we may begin to rule out sexual selection. This will be further addressed by the study outlined below.

In order to test whether the collaborative partner choice mechanism is viable, a study is outlined below. First however, let's briefly test the sexual selection/ mate choice hypothesis using data collected during the study investigating the interplay of cognitive ability and musical learning ability (study two). Alongside measurements of symmetry, cognitive ability and musical learning ability, participants were asked to answer a sexual behaviour questionnaire. This included questions about

1. Sexual orientation (Kinsey et al. (1948) scale, where 0 was exclusively heterosexual, 6 was exclusively homosexual and 7 was asexual)
2. Number of lifetime sexual partners
3. Number of sexual partners in the last 12 months
4. Number of one-time sexual partners (one-night stands)
5. Number of partners the subject has had sex with, without planning to have a long term relationship

The Kinsey scale is adapted from that discussed by Kinsey et al. (1948). The final three questions make up the behavioural facet of the SOI-R (Socio-sexual Orientation Inventory Revised) (Penke and Asendorpf, 2008). According to the sexual selection hypothesis and Miller (2000), we might predict that more intelligent, better musical learners would be those with unrestricted socio-sexuality (more sexual partners). This would speak to the sexual selection hypothesis, as more sexual partners (for men) should theoretically increase potential offspring number in fit (male) individuals. So, if the sexual selection hypothesis is correct, we might expect to see correlations between number of lifetime sexual partners and socio-sexual behaviour and musical learning, cognitive ability, and fluctuating asymmetry.

Cronbach's alpha was calculated for the SOI-R as .87. As this was so reliable, principal component factor analysis was performed to derive one factor, which contributed to 80.64% of variance. This was correlated with the other variables of interest, alongside number of lifetime sexual partners. The non-parametric Spearman's rho correlation coefficient was calculated, as some data (note and time errors) were non-normally distributed. No correlations were found between socio-sexual behaviour (SOI-R) and note errors ($r_s$(100)=-.13, p=.20), time errors ($r_s$(100)=-.11, p=.26), recall number ($r_s$(100)=.11, p=.27), rated creativity ($r_s$(95)=.13, p= .21), rated cultural adherence ($r_s$(95)=-.17, p=.09), g speed ($r_s$(95)=.04, p=.69), or g intelligence ($r_s$(97)=.05, p=.65). Correlations were found between number of lifetime sexual partners and note errors ($r_s$(100)= -.21, p=.03) and recall number ($r_s$(100)=.20, p=.049), but not time errors ($r_s$(100)=-.17, p=.09), rated creativity ($r_s$(95)=.13, p=.22), rated culture ($r_s$(95)=.12, p=.24), g speed ($r_s$(95)=.002, p=.98), or g intelligence ($r_s$(97)=.06, p=.56). If we analyse the data according to gender, we see no correlations between any of the variables of interest and sexual behaviour. This lack of a correlation between sexual behavior and cues to abilities and qualities such as cognitive ability, musical learning ability and fluctuating asymmetry contradicts the predictions of the sexual selection hypothesis, suggesting that sexual selection may not be the sole selection mechanism operating on the proposed LSP. This disputes the theories of Miller (2000), and seems to justify the inclusion of an experiment investigating the role of collaborative partner choice, outlined below.

### 5.1.1 Experimental procedure

This proposed study (study three) had two aims:
1) To test whether perceivers of rhythms truly can pick up on cues of a person's fitness, using only their musical output, and
2) To tease apart the possibly intertwined roles of sexual mate choice and collaborative partner choice in the evolution of this rhythmic precursor.

Raters (N = 49, M = 22) were informed about the format of the original (musical learning) study (study 2), and were told that they would be rating the output of participants in that study. Raters were split into two conditions - sexual and collaborative/social. These conditions can be seen in Table 5.1.

| | Sexual | Collaborative |
|---|---|---|
| Rating | Opposite-Sex (10M) | Opposite-Sex (6M, 7F) |
| Rating | Opposite-Sex (13F) | Same-Sex (6M, 7 F) |

Table 5.1: Conditions for study three, including breakdown of gender of rater

The collaborative condition was further divided into two - opposite and same sex groups;

one group told they were rating the output of same sex participants, the other group told they were rating opposite sex participants. Participants (N = 49) rated 291 rhythms total (one creative (C1) and two replications (R0, R5) each for 97 original participants - five participants from the original experiment did not create their own rhythms), in five experimental blocks. In each block, they first heard the original eight target rhythms from the musical learning study, and then rated 60 rhythms (40 replications, 20 creative) (51 in the final block). In practice, the order of presentation of blocks was randomised, in order to cancel out any order effects. Additionally, order of presentation of rhythms within blocks was randomised. Rating took the form of picking a number on a Likert Scale (1 to 5) to answer the question “How good would this person be as a sexual partner?” or “How good would this person be as a collaborative partner?”, depending on condition. The poles of the Likert Scale were marked “least good” to “very good”. In order to identify attitude towards the opposite sex, only heterosexual participants were recruited for this study. The ratings taken will be referred to as ratings of sexual or collaborative *ability*, rather than *appeal*, as participants were asked to answer how good this person would be as a partner, rather than how likely the rater was to choose them as a partner. Full instructions can be found in the Appendix, on page 163.

The rating took on average one hour and 15 minutes. Two replications and one creative rhythm were selected for rating, from the total output of each original musical learning participant (N=102). These three in particular were chosen, as the replicated rhythms (R0 and 5) were reflective of the differing levels of complexity in the target rhythms, while the creative rhythm (C1 - second creative rhythm made by each participant) was chosen as it was neither the first (perhaps a practice) creative piece, nor the last (when participants may be bored).

We can predict that, if collaborative partner choice is relevant to this system, then ratings should not differ according to sex. This allows us to tease apart the possible roles of social and/or sexual motives in the selection of this melodo-rhythmic precursor. Furthermore, the ratings from this study may be correlated with other variables (processing speed, intelligence, note and time distance, recall number, rated creativity), in order to investigate whether perceivers can pick up on cues in producers’ musical output.

### 5.1.2 ALTERNATIVE APPROACH

The study outlined above takes a very controlled approach to the question in hand. Raters are asked to self-report their opinion on how good the "musician" would be as either a sexual or a collaborative partner. One alternative would be to have the raters engage in a more realistic choice-based game. Rather than rating how good the person might be as a partner, they would actually have to choose a partner from an array of potential partners. They would only be given the musical output of each potential partner from which to make their choice. Then they would have to actually engage in a collaborative task - for example, building a Lego tower. This more realistic approach was taken by Kirschner and Tomasello (2010), who, after having children engage in either joint music making or joint story telling, had these children actually take part in tasks designed to examine their likelihood to engage in a collaborative task or to help a partner. This approach to experimenting is very suitable for use with children, who may not respond well to a simple questionnaire format. However, due to time constraints, and reservations about adequate control conditions, it was decided to take the less realistic, but potentially more easily controlled, experimental approach outlined above.

### 5.1.3 RESULTS

Raters were recruited from the University of Edinburgh psychology undergraduate subject pool and the University of Edinburgh careers website. Psychology undergraduates were given course credit, while recruits from the careers website were paid £7.50 for the on average 75 minute study. Raters (N = 49, M = 22) were split into two conditions, sexual (N = 23, M = 10), and collaborative/ social (N = 26, M = 12). Average age of participants was 20, within a range from 18 to 38. This study was approved by the University of Edinburgh Psychology Department ethics forum. All participants signed consent forms to take part in the study, and for their data to be stored anonymously on a secure server.

In order to investigate whether perceivers would pick up on cues in musical output, the reliability of ratings of each rhythm was calculated. Each rhythm was separated, and the Cronbach's alpha of each rating group calculated. It was found that collapsing across sexual group gave the best reliability for both conditions. The Cronbach's alpha of Collaborative ratings for C1 was .70, for R0 was .79, and for R5 was .81. The Cronbach's alpha of sexual ratings for C1 was .82, for R0 was .74, and for R5 was .80. Each rating group was highly reliable and consistent. These groupings were significantly correlated, condition-internally. Data were found to be largely normally distributed, with collaborative ratings for rhythm 5 non-normal (D(97)=.12, p=.002).

In order to test whether raters in the collaborative condition were rating differently according to whether they believed they were rating opposite or same sex participants, a Mann-Whitney's U test was run on each rhythm (C1, R0, R5). This test was used, as part of the data (collaborative rating of rhythm 5) was non-normally distributed (D(97)=.12, p=.002). No significant difference was found in ratings for C1(U=4411.5, Z=-.75, p=.45) or R5(U=4621.5, Z=-.21, p=.83). Only in R0 was there a significant effect of perceived gender (U=3593.5, Z=-2.84, p=.004). These data, whilst not wholly consistent, indicates that over the three rhythms, raters did not respond differently to same or opposite sex participants. In other words, raters who thought they were rating the collaborative ability of a same-sex participant did not respond differently to raters of an opposite-sex participant. This supports the collapsing of data across genders. The values discussed here can be found in Table 5.2.

| | Opposite | | Same | | Significance |
|---|---|---|---|---|---|
| | Mean | Std. Error | Mean | Std. Error | p value |
| C1 | 3.1 | .03 | 3.0 | .03 | p=.45 |
| R0 | 2.7 | .03 | 2.9 | .03 | p=.004 |
| R5 | 3.0 | .03 | 3.0 | .03 | p=.83 |

Table 5.2: Collaborative ratings for opposite and same sex
This shows no significant overall difference in ratings according to supposed gender of participant

#### 5.1.3.1 Sexual vs Collaborative choice

In order to test whether the groups rating sexual and collaborative ability really were responding differently, a Mann-Whitney's U test was conducted on each rhythm. No significant difference was found between ratings of sexual and collaborative ability for rhythm C1 (U=4214.5, Z=-1.25, p=.21) or rhythm R5 (U=4447, Z=-.66, p=.51). There was a significant difference in ratings for rhythm R0 (U=3363, Z=-3.43, p=.001). These values can be found in Table 5.3. As there is no consistent pattern here, we cannot conclude whether rhythms are being rated differently according to condition. This will be studied further by looking at correlations between these conditions and other variables such as cognitive ability and musical learning ability.

| | Sexual | | Collaborative | | Significance |
|---|---|---|---|---|---|
| | Mean | Std. Error | Mean | Std. Error | p value |
| C1 | 3.1 | .03 | 3.0 | .02 | p=.21 |
| R0 | 2.6 | .02 | 2.8 | .02 | p=.001 |
| R5 | 3.0 | .02 | 3.02 | .02 | p=.51 |

Table 5.3: Mean values for sexual and collaborative ratings

It was found that rated collaborative and sexual ability was highly correlated overall, with correlations between all variables except for between sexual ratings of R0 and collaborative ratings of C1. Details can be found in Table 5.4. Non-parametric Spearman's correlation coefficient was used here, as part of the data was non-normally distributed.

| | Collab. (C1) | Collab. (R0) | Collab. (R5) |
|---|---|---|---|
| Sexual C1 | .81** | .36** | .23* |
| Sexual R0 | .16 | .77** | .29** |
| Sexual R5 | .20* | .34** | .80** |

Table 5.4: Correlations between rated collaborative and sexual ability
*Note*: ** indicates that correlation is significant at 0.01, * indicates that correlation is significant at 0.05

#### 5.1.3.2 RATING AND MUSICAL LEARNING

To test whether raters could pick up on cues to fitness in a producer's musical output, the rated sexual and collaborative quality of each participant was correlated with objective and subjective analyses of their musical learning ability (Note and Time errors, Recall number, rated Creativity and Cultural adherence). Tables below outline the correlations found, using Spearman's Correlation coefficient. The musical learning variables were highly inter-correlated, and principal components factor analysis found that two underlying factors could be extracted from the 5 variables, accounting for 50.20% and 30.94% of the variance respectively. This level of inter-correlation, and multiple comparisons, raise the risk of type 1 error, i.e. of inflating the likelihood of a significant result. For this reason, the alpha level used here was corrected, using Bonferonni's correction, a conservative test which controls the chance of type 1 error rates. As five correlations were conducted between each dependent variable and the musical learning variables, the alpha value of 0.05 should be divided by 5, giving us a corrected alpha of 0.01. This corrected alpha value will be used to identify significant correlations in this analysis.

Table 5.5 shows the correlations between rated collaborative ability, sexual ability and musical learning, rated on the basis of one creative piece (C1) as well as aggregated replications (R0 and R5).

| | Collab. rated C1 | Collab Av. Reps | Sexual rated C1 | Sexual Av. Reps |
|---|---|---|---|---|
| Note Errors | -.27** | -.52** | -.24 | -.33* |
| Time Errors | -.30** | -.62** | -.28** | -.43** |
| Recalled Rhythms | .14 | .51** | .11 | .32** |
| Creativity | .49** | .19 | .67** | .12 |
| Cultural Adherence | .06 | -.01 | -.21 | .01 |

Table 5.5: Correlations between rated collaborative and sexual ability and musical learning
*Note*: ** indicates that correlation is significant at 0.01, the corrected alpha value

This shows that rated collaborative ability correlated with prior rated creativity (for C1 alone), Note and Time errors, and, for the aggregated replications alone, recall number. Rated collaborative ability did not correlate with rated cultural adherence. So, raters seem to be picking up on cues to creativity and ability to copy and later recall pieces correctly, when choosing whether a person would be a good collaborator. There seems to be no role here for cultural adherence, which is perhaps unsurprising in a creative piece of music. Aggregating the rated replications here amplifies the results seen when the replications are considered separately. Rated sexual ability correlated with prior rated creativity (for creative pieces alone), note and time errors, and recall number (for aggregated replications alone). So, raters seem to be picking up on cues to creativity and ability to copy pieces correctly, when choosing whether a person would be a good sexual partner.

In order to draw these results together, we might conclude at this stage that raters seem to be able to pick up on cues of musical learning ability when rating how good a person would be as a collaborative partner. The slightly differing results for creative and replication pieces supports the inclusion of these as separate aspects of the musical learning test - they complement each other as cues of fitness. When rating sexual ability on the basis of creative pieces, raters drew on cues of creativity and replication ability (note and time errors). However, when rating on the basis of replicated pieces, raters drew on different elements of the signal - replication ability and recall ability. These effects were smaller than we have seen in collaborative ratings.

#### 5.1.3.3 Rating and Cognitive Ability

Whilst it is interesting to see that musical learning ability is reliably indicated by musical sequences, the key factor of this thesis is the role of a musical signalling system as a *signal of cognitive ability*. So, correlations were calculated between our measures of cognitive ability (g speed and g intelligence) and rated collaborative and sexual ability. Details can be found in Table 5.6. Correlations were found between g speed and g intelligence and rated collaborative ability, on the basis of aggregated replications. However, no correlations were found between rated collaborative ability on the basis of the creative piece, and either measure of cognitive ability. This indicates that creative pieces are not a good cue of cognitive ability, whereas ability to replicate musical pieces is.

No correlations were found between rated sexual ability and general speed, however, there were correlations between general intelligence and rated sexual ability, on the basis of aggregated replications.

| | Collab. C1 | Collab. Av. Reps | Sexual C1 | Sexual Av. Reps |
|---|---|---|---|---|
| g intelligence | .12 | .44** | .10 | .31** |
| g speed | -.02 | -.33** | .02 | -.13 |

Table 5.6: Correlations between rated collaborative ability, sexual ability and cognitive ability
*Note*: ** indicates that correlation is significant at 0.01

#### 5.1.3.4 Difference between correlations

Whilst we have seen interesting relationships between both musical learning and cognitive ability variables and rated sexual and collaborative ability, it is crucial to test whether there is a significant difference between correlations for sexual and for collaborative ability. So, Fisher's r to Z transformation was conducted on the relationships discussed earlier (Lowry, 2001), and a Z test for correlation differences was run. Contrary to predictions, there was no significant difference in correlation size between rated sexual and collaborative ability, when each were correlated with general intelligence (Z=-1.02, p(2 tailed)=.31) and general speed (Z=1,41, p(2 tailed)=.16) (ratings on the basis of aggregated replications). Furthermore, no significant difference was found between the correlation size of sexual or collaborative ability (on the basis of aggregated replications) and musical learning, note (Z=-1.6, p(2 tailed)=.11), time (Z=-1.82, p(2 tailed)=.07), recall number (Z=1.58, p(2 tailed)=.11), creativity (Z=.49, p(2 tailed)=.62), and cultural adherence (Z=-.01, p(2 tailed)=.89). Thus, whilst we saw a tendency earlier towards different responses according to sexual or social conditions, this was not a statistically significant one.

#### 5.1.3.5 Fluctuating Asymmetry and rating

In order to test the relationship between developmental stability and rated sexual or collaborative skill, Spearman's correlations were calculated between these variables. No correlations were found between any measure of fluctuating asymmetry and the two measures of rated sexual and collaborative ability (C1 and aggregated replications) (See Table 5.7). This is using the Bonferonni-corrected alpha value of 0.00625, due to the number of inter-correlated variables and the resulting risk of type 1 error. However, even prior to correction, no correlations existed. It is perhaps surprising that no correlation exists between rated *sexual* ability and FA, as we may expect sexual partners to find information about genetic fitness relevant. However, the lack of correlation between rated *collaborative* ability and FA supports the hypothesis that collaborative partner choice drove selection, as one would not expect genetic fitness to be relevant in judging the collaborative ability, for a purely non-sexual social task, of a potential partner.

### 5.1.4 Replication of these findings

In a recent MSc project under joint supervision with myself and Simon Kirby, Alex Kiker (2012) conducted a partial replication of the study outlined here. Kiker (2012) altered the paradigm somewhat, by adding a priming element (Griskevicius et al., 2006a,b, 2007, 2009, 2010); raters were exposed to stories designed to elicit sexual, social, or neutral motives. Stories were matched for length and location - each story involved the reader spending an evening at home. The sexual story describes to the reader an evening with a romantic interest, starting with the arrival of his/her date, progressing through conversation, a romantic dinner and a film.

|  | Collab. rated C1 | Collab Av. Reps | Sexual rated C1 | Sexual Av. Reps |
|---|---|---|---|---|
| FA face & body | .08 | -.05 | .03 | -.14 |
| FA face | .16 | -.07 | .14 | -.08 |
| FA body | .11 | .005 | .02 | -.11 |
| FA lower body | .16 | -.06 | .11 | -.17 |
| FA upper body | -.09 | -.05 | -.16 | -.08 |
| FA hands | -.02 | -.13 | -.15 | -.05 |
| FA upper body without hands | -.01 | -.05 | -.06 | -.04 |
| FA body without hands | .13 | -.03 | .04 | -.11 |

Table 5.7: Table showing a lack of correlations between rated collaborative and sexual ability and fluctuating asymmetry

The social story centres around the reader having a social evening with friends. It starts with a cooperative cleaning task, progressing through the communal preparation and eating of dinner, and finishing with a cooperative game. The neutral story was designed to prime only the emotions which may have been primed as by-products of the previous two stories - happiness and contentedness for example. This story involves a solitary, enjoyable evening at home, beginning with the reader cooking dinner, reading a book, then washing the dishes, watching a film and going to bed alone to finish the book. The sexual and social stories included words meant to further prime these motives - words such as “naked”, “stiff”, “skin” for the sexual condition, and “together”, “teamwork”, “community” for the social condition. Analysis found that these stories were successful in eliciting these emotions, with participants in the social condition ending the study with more social feelings than in the sexual or neutral conditions, and so on. This was measured by participant self-report of feelings. Following this priming element, participants were exposed to the original eight rhythms, following the methodology seen in study three, before rating the quality of the replicated rhythms made by participants in study one. Their rating took the form of listening to 40 rhythms (R0 and R5, each replicated by 20 participants), and rating each on a scale from one to seven, with one being a very poor score and seven being a very good one.

Findings mirrored the results detailed above (section 5.1.3), with a larger sample size (N

= 102). It was found that, while socially primed ratings correlated with both g intelligence ($r_s(100)=.47$, $p=.02$) and g speed ($r_s(100)=-.59$, $p < .01$), sexually primed ratings only correlated with g intelligence ($r_s(100)=.45$, $p=.02$). This precisely matches the finding above, and suggests a tendency towards a difference in behaviour between social and sexual ratings. Neutrally primed ratings correlated with both g intelligence ($r_s(100)=.40$, $p=.04$) and g speed ($r_s(100)-.43$, $p=.03$), but to a lesser degree than the socially primed ratings. However, there was no statistically significant difference in the ratings of different groups, according to a Fishers r to Z transformation and subsequent Z test for correlation differences (Lowry, 2001). The non-statistically significant tendency towards different ratings may be amplified by further studies.

## 5.2 DISCUSSION

Teasing apart the different mechanisms of choice in operation on the proposed LSP has proven difficult, and this study has not yielded a perfectly clear-cut picture. Although the results seen above suggest a *tendency* to support the hypothesis of collaborative partner choice, there was no statistically significant difference in correlation size between rated sexual and collaborative ability, here failing to fully support the hypothesis. This suggests that musical learning ability may have acted as a good cue to both the collaborative *and* the sexual abilities of producers of musical signals.

The most interesting aspect of this analysis to date is the correlation between rated collaborative ability and cognitive ability. We found that, when rating the collaborative ability of a potential partner on the basis of aggregated replicated pieces, ratings correlated significantly with g speed and g intelligence. However, when rating the *sexual* ability of potential partners on the basis of aggregated replicated rhythms, whilst a correlation was found between rated sexual ability and g intelligence, the effect size was smaller than that found between collaborative ability and g intelligence. Furthermore, no correlation was found between rated sexual ability and information processing speed. Although this finding is suggestive, there was no statistical difference in correlation sizes between sexual and collaborative conditions.

No relationship emerged between FA and rated sexual or collaborative ability. Whilst the lack of correlation between rated collaborative ability and FA is supportive of the hypothesis that collaborative partner choice drove selection, we cannot conclude firmly in its support here, as there is also a lack of relationship between FA and rated sexual ability.

The experimental results suggest that, in a society in which a learned musical culture was used as a signal of fitness, perceivers could pick cognitively able collaborative partners, on the basis of their musical output. This supports the hypothesis outlined in chapter two. This could have a profound effect on a society, with an honest signalling system being reliably used as a cue to abilities, in people considering a *collaborative* endeavour. However, the data suggests that perceivers could also pick out good potential *sexual* partners, on the basis of musical output. This study may have yielded more robust results with more raters or more rated rhythms, to increase the power. However, at this stage we can conclude that

1) Perceivers of a learned musical signalling system could choose cognitively able collaborative partners, based solely on musical output.

2) Perceivers could also use these musical signals as a cue to the sexual fitness of a potential mate.

These results have an impact upon the question of selection in language evolution - we need not focus on sexual selection or mate choice as a sole mechanism for the evolution of signalling systems to language, but rather, consider other mechanisms, such as collaborative partner choice. Future studies may further pull apart the perhaps richly interwoven roles of mate and collaborative partner choice in the evolution of this learned musical precursor.

The most striking result from this study is the fact that raters were able to pick up on both musical ability (as measured both objectively and subjectively) *and* cognitive ability (measured objectively), based solely on very short, low bandwidth signals. Raters had only a few seconds to listen to each rhythm, created using just two tones, in a sterile and relatively unusual environment (that of the experiment), and yet they were able to use cues in the musical output of original participants to inform their decisions about mating or collaboration potential. This is not a trivial task. This is interesting, and suggestive that these raters may have some sensitivity to pick up on cues such as those held within the output rhythms. Perhaps this is indicative of a human propensity to respond to cues theoretically held within a musical signal.

# CHAPTER 6

# CONCLUSION AND IMPLICATIONS

This thesis adds to the growing literature on musical protolanguage theory, a field which has seen a revival in recent decades. First proposed by Darwin (1871a), the idea that language may have evolved from music has been developed, at times regrettably without adequate reference to Darwin himself. Now, just over 140 years since its inception, seems an appropriate time for empiricists to add to the theoretical discussion of music's role in language origins. In this thesis, I have laid out a framework for the evolution of language from a musical precursor. This was a learned sequential precursor, which underwent social selection driven by collaborative partner choice. Further to that, I have tested the predictions of the hypothesis, in three studies summarised below.

The main aim of this thesis is to provoke thought about testing musical protolanguage theories. In generating testable hypotheses, we must think about the pressures that acted on a proposed system, how it functioned in social groups, where it originated and how it evolved. Thinking separately about each of these avenues will hopefully allow the testing of pointed hypotheses, and I believe that this thesis has addressed each of these concerns. This thesis scaffolds a series of empirical tests which study the function of the proposed learned sequential precursor, analysing some traits of the precursor, and attempting to shed light on the selection mechanisms this system may have undergone. This will hopefully set the groundwork for a rich empirical investigation into the origins and evolution of language from a musical precursor.

## 6.1 Summary of Study One

The first study looked in detail at the perceived complexity of pieces of music. Is perceived ability to create complexity in music associated with other traits? Participants were asked to listen to complex and non-complex pieces of music (as measured by number of rhythmic elements), and to rate each according to how much they liked the piece, how familiar the rhythm was, how likely they were to pick the player to do a tricky treasure hunt with, how attractive and intelligent they thought the person was. In reality, they were not rating real pieces of music as made by a person, but computer-synthesised pieces.

A significant correlation was found between complexity level and all questions excluding familiarity and collaboration, with participants preferring more complex pieces, and finding these "musicians" more intelligent and attractive. The question regarding intelligence indicated that complexity in rhythm may act as a signal of cognitive ability, in choosing a collaborative partner. Overall, this study suggests that complexity in musical rhythms may have been valued as a signal, and that perhaps this preference for complexity remains in modern populations as a remnant of the prior complex musical sequences, which were used as signals of cognitive ability.

## 6.2 Summary of Study Two

The main study of this thesis aimed to test whether musical ability really could signal cognitive ability and developmental stability. The LSP hypothesis proposes that a pre-linguistic system consisted of a musical signalling system, the learned sequential precursor or LSP, which acted as a cue to the cognitive ability of producers, as it was honestly linked to their ability to learn a complex cultural system. This thesis further proposes that social selection was key to the evolution of the LSP.

Participants underwent an intelligence test (with verbal, numerical and perceptual aspects), an information processing speed test, a musical learning test, and had the fluctuating asymmetry of their bodies, faces and hands measured. It was predicted that, if the social selection theory was upheld, that cognitive ability (as measured by general intelligence and processing speed) and musical learning ability would correlate. Fluctuating asymmetry (a predictor of developmental stability) was not predicted to correlate with the other variables under the social selection hypothesis. This study took a novel approach to musical protolanguage theories, employing an entirely new musical learning test, which was proven to be reliable and consistent.

It was found that cognitive ability did indeed correlate with musical learning ability, with both measures of cognitive ability - general intelligence and processing speed - correlating significantly with two of the measures of learning, ability to replicate and numbers of rhythms recalled. However, the third measure of musical learning, creativity, did not correlate significantly with cognitive ability. In other words, more intelligent people who react more quickly to experimental stimuli, are also better able to copy musical pieces and to later recall them, but not necessarily to create them. This upholds the prediction that musical ability could signal cognitive ability. Creativity here may not be an appropriate measure of musical learning ability, however it is interesting to note the negative correlation found between creativity and cultural adherence. It was found that one cannot be both creative and culturally adherent, but if one is rated low on cultural adherence, there is still a range of creativity open to them.

No correlations were found between Fluctuating Asymmetry and cognitive ability measures or between FA and musical learning measures. This is expected, given the social selection hypothesis, as, for a purely *social* endeavour, genetic robustness should be irrelevant. Genetic quality (as indirectly measured here by FA) should not be relevant to individuals operating in a social system, in which no mating is expected.

## 6.3 SUMMARY OF STUDY THREE

This study aimed to tease apart the roles of mate and collaborative partner choice in the selection of the musical signalling system. I have proposed collaborative partner choice driving social selection as a key mechanism in the evolution of this system - the musical signalling system acted as a reliable indicator of the abilities of each potential collaborator in a population. This allowed perceivers to pick cognitively able individuals to collaborate with, leading to increased resources for both signaller and perceiver. This meant that signallers, (who had the ability to learn a novel complex cultural system) would have lived longer and been more healthy, thereby increasing their potential reproductive fitness. This ability to learn would have passed to the next generation, allowing it to spread throughout the population.

This study also tested whether perceivers truly could pick up on cues of abilities in a musical signal. Raters listened to rhythms created by participants in study one, and rated either the sexual or collaborative ability of the participants, on the basis of their musical output (a mix of creative and reproduced pieces). This test was reliable and consistent. When rating the collaborative ability of participants, correlations were found between rated ability and creativity, replication ability and number of rhythms recalled. When rating sexual ability, correlations were found between rated ability and creativity, replication ability, and number of rhythms recalled. Correlations were found between rated collaborative ability and both measures of cognitive ability, general intelligence and general processing speed. This pattern on first view appears to be different to that seen when comparing rated *sexual* ability and cognitive ability. Here, only general intelligence correlated with rated sexual ability. However, this was a non-significant difference. In this study, the greatest indicator of cognitive ability appear to be the replicated pieces of music, and when rating sexual ability, processing speed is not a good indicator. No correlations were found between FA and either rated collaborative ability or sexual ability.

This study leaves open some questions. While raters did seem to behave slightly differently when rating sexual and collaborative ability, seeming to draw on different cues in the musical signals, this difference was not statistically significant. Future expansion of this study may accentuate the tendencies seen here. This study goes some way towards addressing the bias towards sexual selection and mate choice in the musical protolanguage literature, by suggesting collaborative partner choice and social selection as a viable alternative.

As seen above, the most interesting result of this experiment was that raters were actually able to make informed decisions about the ability of original participants, based solely on short, sanitized, low bandwidth signals. Their subjective measures of sexual or collaborative ability correlated with both objective and subjective measures of cognitive ability and musical leaning ability.

## 6.4 Implications of Experimental Paradigm

The studies outlined in this thesis lend substantial support to the hypothesis suggested in chapter two, of a complex, musical, learned sequential signalling system as a precursor to language. The first study indicated that humans do have a preference for complexity in musical rhythms, supporting the view that the LSP consisted of *complex* learned sequences. The second provided evidence that musical learning ability is correlated with cognitive ability, and the third suggested that different selection pressures (sexual and social) may have been in operation on this system.

The three studies on which this thesis is based together support the hypothesis that complex, culturally learned musical signals could have acted as an honest signal of cognitive ability, in a group which used such a culturally learned system. An initial stage in the evolution of a musical precursor to language may have taken the form of a musical group coordination system. This may have been scaffolded by a mother-child interaction system, and may have been used by groups to maintain social bonds and commitments. From this stage, biological evolution would have been driven by the pressure for an honest signalling system of cognitive ability. This was facilitated by the social and environmental changes associated with the advent of Homo erectus. This led to a system of signalling cognitive ability through the use of complex, learned, musical sequences which were facilitated by membership of a cultural group and an investment in learning. The learned sequential precursor or LSP did not have semantic meaning. This system may not have changed or evolved further until the emergence of Homo sapiens. With Homo sapiens came a pressure both for expressivity and semantic meaning. This led to a protolanguage which may have been musical in nature and was holistic and semantically meaningful. Cultural evolution, plus the competing pressures of expressivity and learnability would have led to the segmentation of this system (as we have seen in experiments and models by Kirby (2001); Kirby et al. (2008); Verhoef et al. (2011)), leading in turn to a compositional, semantically meaningful, complex, learned system - language. From the musical learned sequential precursor, music is proposed to have evolved, perhaps through a protomusical stage, with the addition of emotional meaning creating a non-referential, complex, emotionally meaningful system, stripped of referential meaning.

### 6.4.1 RE-ANALYSIS OF THE STUDY OF FA

This thesis has revealed some interesting implications for the study of Fluctuating Asymmetry. Previous studies (Furlow et al., 1997; Bates, 2007; Prokosch et al., 2005; Penke et al., 2009), have found correlations between g intelligence, cognitive aging, and FA. Indeed, a rich and growing field has grown around examining relationships between FA and many other aspects of life, such as personality (Hope et al., 2010), and FA is seen as a useful inverse predictor of developmental stability. However, this thesis failed to replicate negative correlations between

FA and cognitive ability. Further to that, a recent high-profile paper claiming relationships between symmetry and rated dancing ability (Brown et al., 2005) has been roundly criticised, by a group of authors which includes one of the co-authors of the original paper (Trivers et al., 2009). This book states that a thorough re-analysis of Brown et al. (2005) reveals that all the major results appear to have been based on hidden procedures designed to produce the results found. This consisted in part of pre-selection participants based on dancing ability, and also in incorrectly averaging dance ability scores in order to create significant results where none existed. This case suggests that we may need to be more cautious in our conclusions about FA, and our interpretation of results. We need to question whether the group we are using is likely to have actually undergone developmental instability, and what conditions we might identify FA under. Is the use of Western, Educated, Industrialised, Rich, Democratic (WEIRD) societies valid? This question may be answered by noting that the vast majority of psychology studies are conducted on exactly that population; however, this is certainly an issue which needs to be addressed if we are to conclude universal behavioural predictions based on these data. Future research should aim to address the conditions under which FA measures are useful, and whether using undergraduate students to inform global assumptions about developmental stability and fitness is appropriate.

### 6.4.2 USE OF MUSIC AS A PROXY FOR LANGUAGE

Discussion of evolutionary pathways shared between music and language seems to justify the use of music as a proxy for language in experiments into the origins of language. For many reasons such as first language transfer effects, it is sometimes inappropriate to use language-based stimuli in an experiment exploring people's use of language, or investigating its evolution, for example through iterated learning. Use of music as a proxy for language is seen in the work of Verhoef et al. (2011, 2012), who used slide whistles as a proxy for spoken language, in order to abstract away from compositional structure and use a continuous signal space. Tamariz et al. (2010) also used musical tones as a proxy for language, in order to investigate the role of literacy and practice in the evolution of linguistic structure. This use of music as a proxy has been further supported by this thesis, as the experimental investigation implies a shared predecessor of both music and language.

### 6.4.3 WAYS FORWARD FOR MUSICAL PROTOLANGUAGE THEORIES

This thesis has proposed a robust empirical test of musical protolanguage theories. However, much is still to be done. Some of the work presented here was necessarily restricted in scope, with limited time and finances restricting, for example, numbers of participants. A replication of the selection study with more participants may contribute to the understanding of the roles of social and sexual selection in the evolution of the LSP.

The theory that music played some role in the evolution of language must now, with nascent empirical work, be accepted as a useful, workable theory in the language evolution field. For too long has this theory been scoffed at and relegated to an untestable just-so story (e.g Weiss (1974)). The experimental work laid out in this thesis has proven that hypotheses can be tested, if we draw on disparate literatures. If we are to address the many questions left open by theories of the evolution of language, we must do so through drawing on all available options - mathematical modelling, computer simulations, as well as empirical investigation with modern humans and extant apes. This empirical investigation will hopefully continue apace, with this thesis, and ongoing work from research groups such as the Language Evolution and Computation Research unit in Edinburgh and the Cognitive Biology department in Vienna, adding to the growing theoretical interest. I trust that this thesis has made a meaningful contribution to the field.

# Appendix A

## Appendix of abbreviations

| | |
|---|---|
| LSP | Learned Sequential Precursor |
| FA | Fluctuating Asymmetry |
| DS | Developmental Stability |
| DI | Developmental Instability |
| ESS | Evolutionarily Stable Strategy |

# APPENDIX B

# APPENDIX

## B.1 STUDY ONE - INSTRUCTIONS

Page 1: “You are an astronaut, sent to the planet Zog to collect data about their musical culture. These aliens are interesting as they communicate using music. You will now hear some examples of the music the aliens on Zog create. Your job is to answer questions about the music and the musician who created it”.

Page 2: “Please pick a number between 1 and 7, according to what you think of the music or musician”.

After page 2, pages cycled through, matching each of 8 rhythms against the following 5 questions:
“How much do you like this rhythm? 1 = not at all, 7 = very much”
“How familiar is this rhythm? 1= not at all, 7 = very”
“How likely are you to pick this person to do a tricky treasure hunt with? 1= not at all, 7 = very”
“How attractive do you think this musician is? 1 = not at all, 7 = very”
“How intelligent do you think this musician is? 1 = not at all, 7 = very”

Final Page: “ Thank you very much for your assistance. Please tell the experimenter that you have finished. Click now to end the experiment”.

## B.2 Study Two - Instructions

Page 1: “Welcome to this game. Today, you will be learning some alien music. There are clear instructions throughout. Press the GO button to proceed”.

Page 2: “This is your chance to get familiar with the musical instrument in front of you. Play with the buttons circled in the picture - the ones on the right are low, and on the left are high. You might have to hit the buttons quite hard to make sounds. When you want to proceed, press the Continue button. Throughout this game, you should only ever press these buttons”.

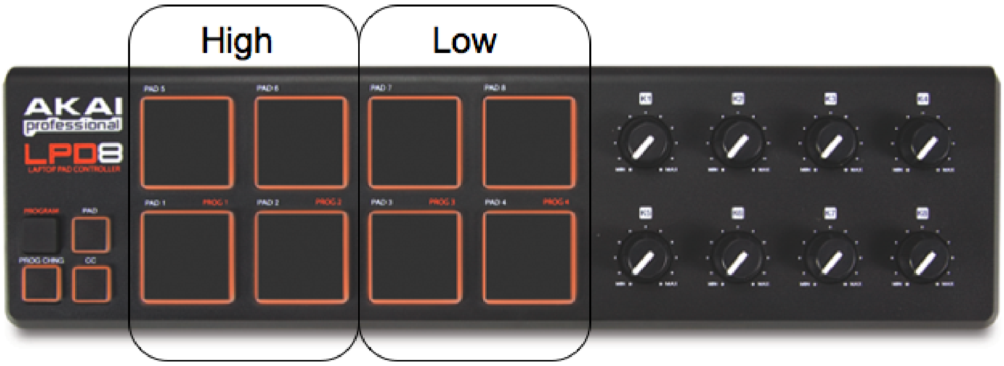


Figure B.1: Picture presented to participants, showing the “high and low buttons” on the novel instrument

Page 3: “In this section of the game, you will hear a steady rhythm. Try to tap along to it while it plays. At some point, it will stop playing, but please keep tapping the rhythm (like alien karaoke!) until the “Start” button is replaced by a red “Done” button below. Press this button to proceed”.

Page 4: “In the next section of the game, you will hear some musical rhythms. There are 8 rhythms, each more difficult than the last. You will get 4 attempts to copy each - at each attempt you will hear the rhythm and have a chance to copy it”.

Page 5: (Rhythm plays) “Only start copying when the rhythm is finished. Press the Next button when you have finished playing each time”. (A progress bar filled up as participants cycled through the 32 playings of rhythms).

Page 6: “You have finished the copying section of the game. In the next section, you must try to recall the 8 rhythms you just heard. The order you remember them doesn’t matter. Press the Recall button to continue.

Page 7: “Using the instrument in front of you, play the rhythms you remember. After each one, press the “Submit” button”.

Page 8: “In this section, you get to be creative! Try to make 4 new rhythms that sound somewhat like the others you just copied, whilst still being different. Press the Create button to continue”.

Page 9: “Press the Enter button when you have finished playing a new piece”.

Page 10: “Congratulations! You have completed the game. Please let the experimenter know that you’re finished”.

## B.3 Edit Distance program

```python
#!/usr/bin/env python

from math import *
import sys

def timingdist(s1, s2):
    if len(s1) < len(s2):
        return timingdist(s2, s1)
    if not s1:
        return sum(s2)
    previous_row=[0]
    for c in s2:
        previous_row.append(previous_row[-1]+c)
    for i, c1 in enumerate(s1):
        current_row = [previous_row[0]+c1]
        for j, c2 in enumerate(s2):
            insertions = previous_row[j + 1] + c1
            deletions = current_row[j] + c2
            substitutions = previous_row[j] + abs(c1-c2)
            current_row.append(min(insertions, deletions, substitutions))
        previous_row = current_row
    return previous_row[-1]

def notedist(s1, s2):
    if len(s1) < len(s2):
        return notedist(s2, s1)
    if not s1:
        return len(s2)
    previous_row = xrange(len(s2) + 1)
    for i, c1 in enumerate(s1):
        current_row = [i + 1]
        for j, c2 in enumerate(s2):
            insertions = previous_row[j + 1] + 1
            deletions = current_row[j] + 1
            substitutions = previous_row[j] + (c1 != c2)
            current_row.append(min(insertions, deletions, substitutions))
        previous_row = current_row
    return previous_row[-1]

def processfiles(f1,f2):
    d1=[l.rstrip().split(',') for l in open(f1).readlines() if l.strip()]
    d2=[l.rstrip().split(',') for l in open(f2).readlines() if l.strip()]
    return notedist([l[0] for l in d1],[l[0] for l in d2]), timingdist([float(l[1]) for l in d1],[float(l[1]) for l in d2])

results=processfiles(sys.argv[1],sys.argv[2])
print "%i,%f" % results
```

Figure B.2: Edit Distance program used for analysis of Study One, written by Prof. Simon Kirby

## B.4 STUDY THREE - INSTRUCTIONS

Page 1: “You will be rating pieces of music made by participants of another experiment. They listened to 8 pieces of music, and were then asked to copy these pieces of music. Additionally, they were asked to make creative pieces of music. Next, you will hear the 8 pieces of music each participant heard”.

Page 2: “Press the button labelled rhythm 1 to hear the first rhythm, and so one. When you have listened to them all, press the Proceed button”.

Page 3: “Now you have heard the 8 rhythms each participant heard. Remember that each participant was asked to copy these rhythms and to make something creative, based on the rhythms they’d heard. Now you will rate each participant based on the rhythms they created”.

Page 4 (Collaborative/Social condition): “You will now hear rhythms made by 20 participants. Some are copies of the rhythms they heard, and some are their creative output. Your job is to rate each participant as a collaborative partner: How competent would this person be at working with you on a tricky task?”.

Page 4 (Sexual condition): “You will now hear rhythms made by 20 participants. Some are copies of the rhythms they heard, and some are their creative output. Your job is to rate each participant as a sexual partner: How likely are you to choose them to have a one night stand with?”.

Page 5 (Collaborative/Social condition): “Please listen to this rhythm and select a number between 1 and 5. 1 is least good as a collaborative partner and 5 is very good. Press enter to continue”.

Page 5 (Sexual condition): “Please listen to this rhythm and select a number between 1 and 5. 1 is least good as a sexual partner and 5 is very good. Press enter to continue”.

Page 6: “Congratulations! You have completed the rating experiment. Please let the experimenter know that you’re finished”.

# APPENDIX C

# LEGEND FOR ATTACHED CD

The attached CD contains the stimulus rhythms for the first and second studies. These rhythms correspond with Figures 4.1, 4.2, and 4.4.

1. Study One Complex Rhythm 5
2. Study One Complex Rhythm 6
3. Study One Complex Rhythm 7
4. Study One Complex Rhythm 8
5. Study One Noncomplex Rhythm 1
6. Study One Noncomplex Rhythm 2
7. Study One Noncomplex Rhythm 3
8. Study One Noncomplex Rhythm 4
9. Study Two Rhythm 0
10. Study Two Rhythm 1
11. Study Two Rhythm 2
12. Study Two Rhythm 3
13. Study Two Rhythm 4
14. Study Two Rhythm 5
15. Study Two Rhythm 6

16. Study Two Rhythm 7

# BIBLIOGRAPHY

Adorno, T. (1956). *Quasi Unia Fantasia: Essays on Modern Music*. R. Livingstone (transl). Verso, London/New York.

Aiello, L. and Dunbar, R. (1993). Neocortex size, group size and the evolution of language. *Current Anthropology*, 34(2):184 – 193.

Albert, F., Schepina, O., Winter, C., Römpler, H., Teupser, D., Palme, R., ..., and Pääbo, S. (2008). Phenotypic differences in behaviour, physiology and neurochemistry between rats selected for tameness and for defensive agression toward humans. *Hormones and Behaviour*, 53(3):413 – 421.

Alexander, G., Swerdloff, R., Wang, C., Davidson, T., McDonald, V., Steiner, B., and Hines, M. (1997). Androgen-behaviour correlates in hypogonadal men and eugonadal men. *Hormones and Behaviour*, 31(2):110 – 119.

Amabile, T. (1983a). The social psychology of creativity: A componential conceptualisation. *Personality Processes and Individual Differences*, 45(2):357 – 376.

Amabile, T. M. (1983b). *The social psychology of creativity*. Springer Verlag, New York.

Arbib, M. (2003). The evolving mirror system: A neural basis for language readiness. In Christiansen, M. and Kirby, S., editors, *Language Evolution*, pages 182 – 200. Oxford University Press.

Arcadi, A., Robert, D., and Boesch, C. (1998). Buttress drumming by wild chimpanzees: Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates*, 39(4):505 – 518.

Arcadi, A., Robert, D., and Mugurusi, F. (2004). A comparison of buttress drumming by male chimpanzees from two populations. *Primates*, 45(2):135 – 139.

Axelrod, R. and Hamilton, W. (1981). The evolution of cooperation. *Science*, 211(4489):1390 – 1396.

Banks, G., Batchelor, J., and McDaniel, M. (2010). Smarter people are (a bit) more symmetrical: A meta-analysis of the relationship between intelligence and fluctuating asymmetry. *Intelligence*, 38(4):393 – 401.

Baron-Cohen, S., Scott, F. J., Williams, J., Bolton, P., Matthews, F. E., and Brayne, C. (2009). Prevalence of autism-spectrum conditions: UK school-based population study. *British Journal of Psychiatry*, 194(6):500 – 509.

Baroni, M., Maguire, S., and Drabkin, W. (1983). The concept of musical grammar. *Music analysis*, 2(2):175 – 208.

Barrett, L., Dunbar, R., and Lycett, J. (2002). *Human Evolutionary Psychology*. Princeton University Press, Princeton, N.J.

Bartlett, F. (1932). *On Remembering*. Macmillan, Oxford.

Bates, T. (2007). Fluctuating asymmetry and intelligence. *Intelligence*, 35(1):41 – 46.

Beauvois, M. (2007). Quantifying aesthetic preference and perceived complexity for fractal melodies. *Music Perception*, 24(3):247 – 264.

Bennett, M., Schatz, M. F., Rockwood, H., and Wiesenfeld, K. (2002). Huygen's clocks. *Proceedings of the Royal Society of London - Series A*, 458(2019):563 – 579.

Berlyne, D. (1958). The influence of complexity and novelty in visual figures on orienting responses. *Journal of Experimental Psychology*, 55(3):289 – 296.

Bickerton, D. (1981). *Roots of Language*. Karoma Publishers, University of California.

Bickerton, D. (1990). *Language and Species*. University of Chicago Press, Chicago.

Bickerton, D. (1996). *Language and Human Behaviour*. UCL Press, London.

Bickerton, D. (2003). Symbol and structure: A comprehensive framework for language evolution. In Christiansen, M. and Kirby, S., editors, *Language Evolution*, pages 77 – 93. Oxford University Press, Oxford.

Bigand, E., Perruchat, P., and Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Current Pychology of Cognition*, 17(3):577 – 600.

Bipsham, J. (2009). Music's "design features" - musical motivation, musical pulse and musical pitch. *Musicae Scientiae Special Issue on Music and Evolution*, 13(2):29 – 44.

Blount, B. (1984). Mother-infant interaction: features and functions of parental speech on English and Spanish. In Pellegrini, A. and Yawkey, T., editors, *The development of oral and written language in social contexts*. Ablex, New Jersey.

Bobe, R. and Leakey, M. (2009). Ecology of the Plio-Pleistocene mammals in the Omo-Turkana basic and the emergence of Homo. In Grine, F., Fleagle, J., and Leakey, R., editors, *The first humans - origins and early evolution of the genus Homo*, pages 173 – 184. Springer, Berlin.

Bolhuis, J., Okanoya, K., and Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, 11(11):747 – 759.

Boogert, N., Giraldeau, L., and Lefebre, L. (2008). Song complexity correlates with learning ability in zebra finch males. *Animal behaviour*, 76(5):1735 – 1741.

Borchgrevink, H. (1991). Prosody, musical rhythm, tone pitch, and response initiation during amytal hemisphere anaesthesia. In Sundberg, J., Nord, L., and Carlson, R., editors, *Music, Language, Speech and Brain*, pages 327 – 343. Macmillan, Houndmills, UK.

Bornstein, M. H. (2006). Parenting science and practice. In Renninger, K. and Sigel, I., editors, *Handbook of Child Psychology: Child Psychology in Practice*, volume four, pages 893 – 949. Wiley and Sons, 6th edition.

Boyd, R. and Richerson, P. (1990). Group selection among alternative evolutionarily stable strategies. *Journal of Theoretical Biology*, 145:331 – 342.

Brown, S. (2000). The "musilanguage" model of music evolution. In Wallin, N., Merker, B., and Brown, S., editors, *The origins of music*, pages 271 – 300. MIT Press, Massachusetts.

Brown, W., Price, M., Kang, J., Pound, N., Zhao, Y., and Yu, H. (2008). Fluctuating asymmetry and preference for sex-typical body characteristics. *Proceedings of the National Academy of Sciences*, 105(35):12938 – 12943.

Brown, W., Vronk, L., Grochow, K., Jacobson, A., Lui, C., Popovic, Z., and Trivers, R. (2005). Dance reveals symmetry especially in young men. *Nature*, 438(7071):1148 – 1150.

Bruner, J. (1962). The conditions of creativity. In Gruber, H., editor, *Contemporary approaches to creative thinking*. Atherton Press, New York.

Burnett, J. (1774). *Of the origin and progress of language*. Balfour, Edinburgh, 2nd edition.

Byrne, R. and Whiten, A. (1988). *Machiavellian Intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*. Clarendon Press, Oxford.

Capasso, L., Michetti, E., and D'Anastasio, R. (2008). A *Homo erectus* hyoid bone: Possible implications for the origin of the human capability for speech. *Collegium Anthropologicum*, 32(4):1007 – 1011.

Carter, S. (1992). Oxytocin and sexual behavior. *Neuroscience and Behavioural Review*, 16(2):131 – 144.

Catchpole, C. and Slater, P. (1995). *Birdsong: Biological themes and variation*. Cambridge University Press, Cambridge.

Charlton, B. D., Filippi, P., and Fitch, W. T. (2012). Do women prefer more complex music around ovulation? *PLoS One*, 7(4):1 – 6.

Claes, P., Walters, M., Vandermeulen, D., and Clement, J. (2011). Spatially-dense 3D facial symmetry assessment in both typical and disordered growth. *Journal of Anatomy*, 219(4):444 – 455.

Conley, J. (1981). A psychophysical investigation of judgments of complexity in music. *Psychomusicology: A journal of research in music cognition*, 1(2):59 – 71.

Cramer, D. (1997). *Badic statistics for social research: Step-by-step calculations and computer techniques using minitab*. Routledge, London.

Crook, J. (1972). Sexual selection, dimorphism and social organisation in the primates. In Campbell, B., editor, *Sexual selection and the Descent of Man: The Darwinian Pivot*, pages 231 – 281. Aldine, Chicago.

Cross, I. (1999). Is music the most important thing we ever did? Music, development and education. In Yi, S. W., editor, *Music, mind and science*, pages 10 – 39. Seoul National University Press.

Cross, I. (2003). Music and evolution: Consequences and causes. *Contemporary Music Review*, 22(3):79 – 89.

Cross, I. (2009). The evolutionary nature of musical meaning. *Musicae Scientiae Special issue on Music and Evolution*, 13(2):179 – 200.

Darwin, C. (1871a). *The descent of man and selection in relation to sex*, volume 1. John Murray, London, 1st edition.

Darwin, C. (1871b). *The descent of man and selection in relation to sex*, volume 2. John Murray, London, 1st edition.

Darwin, C. (1874). *The descent of man and selection in relation to sex*. John Murray, London, 2nd edition.

Deacon, T. (1997). *The symbolic species*. WW. Norton and Company Inc, New York.

Deacon, T. (2003). Multilevel selection in a complex adaptive system: The problem of language origins. In Weber, B. and Depew, D., editors, *Evolution and learning; the Baldwin effect reconsidered*, pages 81 – 106. MIT Press, Massachusetts.

Deary, I., Liewald, D., and Nissan, J. (2011). A free, easy-to-use, computer-based simple and four-choice reaction time programme: The Deary-Liewald reaction time task. *Behavioral Research*, 43(1):258 – 268.

Dion, K., Berscheid, E., and Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3):285 – 290.

Dissanayake, E. (2000). Antecedents of the temporal arts on early mother-infant interaction. In Wallin, N., Merker, B., and Brown, S., editors, *The origins of music*, pages 389 – 410. MIT Press.

Dixon, J. (1979). Quality versus quantity: The need to control for the fluency factor in originality scores from the Torrance tests. *Journal for the Education of the Gifted*, 2(2):70 – 79.

Dongen, S. (2006). Fluctuating asymmetry and developmental instability in evolutionary biology: past present and future. *European Society for Evolutionary Biology Journal Compilation*, 19(6):1727 – 1743.

Doupe, A. and Kuhl, P. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Reviews of Neuroscience*, 22(1):567 – 631.

Dunbar, R. (1984). *Reproductive decisions: An economic analysis of the social strategies of Gelada babboons*. Princeton University Press, Princeton.

Dunbar, R. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6):469 – 493.

Dunbar, R. (2003). The social brain: mind, language and society in evolutionary perspective. *Annual Review of Anthropology*, 32:163 – 181.

Dunbar, R. (2004). Language, music and laughter in evolutionary perspective. In Kimbrough Oller, D. and Griebel, U., editors, *Evolution of communication systems: A comparative approach*. MIT Press, Massachusetts.

Dunsworth, H. (2010). Origin of the genus Homo. *Evolution: Education and Outreach*, 3(3):353 – 366.

Erdi, P. (2008). *Complexity explained*. Springer Verlag, Berlin.

Falk, D. (2004). Prelinguisic evolution in hominin mothers and babies: Whence motherese? *Behavioural and Brain Sciences*, 27(4):491 – 503.

Fantz, R. (1961). The origin of form perception. *Scientific American*, 204(5):66 – 72.

Federenko, E., Oatel, A., Casasanto, D., Winawer, J., and Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory and cognition*, 37(1):1–9.

Fenk-Olczon, G. and Fenk, A. (2009). Some parallels between language and music from an evolutionary perspective. *Musicae Scientiae Special Issue on Music and Evolution*, 13(2):201 – 226.

Ferdinand, V. and Zuidema, W. (2009). Thomas' theorem meets Bayes' rule: a model of the iterated learning of language. In Taatgen, N. and van Rijn, H., editors, *Proceedings of the 31st*

*Annual Conference of the Cognitive Science Society*, pages 1786 – 1791. Cognitive Science Society.

Fink, B. (2002). The Neanderthal flute and origin of the scale: fang or flint? A response. In Hickmann, E., Kilmer, A. D., and Eichmann, R., editors, *Studies in Music Archaeology III*, pages 83 – 87. Verlag Marie Leidorf GmbH, Berlin.

Finlay, B. and Darlington, R. (1995). Linked regularities in the development and evolution of mammalian brains. *Science*, 268(5217):1578 – 1584.

Finnas, L. (1989). How can musical preferences be modified? *Bulletin of the Council for Research in Music Education*, 102:1 – 58.

Fitch, W. T. (2005a). The evolution of music in comparative perspective. *Annals of the New York Academy of Sciences*, 1060:1–20.

Fitch, W. T. (2005b). Protomusic and protolanguage as alternatives to protosign. *Behavioural and Brain Sciences*, 28(2):132 – 133.

Fitch, W. T. (2006a). The biology and evolution of music: a comparative perspective. *Cognition*, 100(1):173 – 215.

Fitch, W. T. (2006b). On the biology and evolution of music. *Music Perception*, 24(1):85 – 88.

Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press, Cambridge.

Fitch, W. T. (2012). The biology and evolution of rhythm: Unravelling a paradox. In Rebuschat, P., Rohrmeier, M., Hawkins, J., and Cross, I., editors, *Language and music as cognitive systems*, pages 73 – 95. Oxford University Press, Oxford.

Fitch, W. T. and Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Perception*, 25(1):45 – 58.

Freitag, C. M. (2007). The genetics of autistic disorders and its clinical relevance: a review of the literature. *Molecular Psychiatry*, 12(1):2 – 22.

Fukui, H. (1998). The effect of music listening on testosterone secretion. *Musical Perception and Cognition*, 4:10 – 18.

Fukui, H. (2001). Music and testosterone: A new hypothesis for the origin and function of music. *Annals of the New York Academy of sciences*, 930:448 – 451.

Fukui, H. and Yamashita, M. (2003). The effects of music and visual stress on testosterone and cortisol in men and women. *Neuroendocrinology Letters*, 24(3-4):173 – 180.

Furlow, B., Armijo-Prewitt, T., Gangestad, S., and Thornhill, R. (1997). Fluctuating asymmetry and psychometric intelligence. *Proceedings of the Royal Society of London - Series B: Biological Sciences*, 264(1383):823 – 829.

Furlow, B., Gangestad, S., and Armijo-Prewitt, T. (1998). Developmental stability and human violence. *Proceedings of the Royal Society of London - Series B: Biological Sciences*, 265(1390):1 – 6.

Gariépy, J., Bauer, D., and Cairns, R. (2001). Selective breeding for differential agression in mice provides evidence for heterochrony in social behaviours. *Animal Behaviour*, 61(5):933 – 947.

Gell-Mann, M. (2003). Regularities and randomness: Evolving schemata in science and the arts. In Casti, J. and Karlqvist, A., editors, *Art and Complexity*, pages 47 – 58. Elsevier Science, Amsterdam.

Getty, T. (2002). Signaling health versus parasites. *The American Naturist*, 159(4):363 – 371.

Gioia, T. (2006). *Work songs*. Duke University Press, London.

Graham, J., Emlen, J., Freeman, D., Leamy, L., and Kieser, J. (1998). Directional asymmetry and the measurement of developmental instability. *Biological Journal of the Linnean Society*, 64(1):1–16.

Gray, A. (1876). *Darwiniana: Essays and Reviews Pertaining to Darwinism*. Cambridge University Press, Cambridge.

Greenfield, M. D. (1994). Cooperation and conflict in the evolution of signal interactions. *Annual Review of Ecology and Systematics*, 25:97 – 126.

Griffiths, T. and Kalish, M. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, 31(3):441 – 480.

Griskevicius, V., Cialdini, R., and Kenrick, D. (2006a). Peacocks, picasso, and parental investment: The effects of romantic motives on creativity. *Journal of Personality and Social Psychology*, 91(1):63 – 76.

Griskevicius, V., Goldstein, N., Mortensen, C., Gialdinn, R., and Kenrick, D. (2006b). Going along versus going alone: When fundamental motives facilitate strategic (non)conformity. *Journal of Personality and Social Psychology*, 91(2):281 – 294.

Griskevicius, V., Shiota, M., and Nowlis, S. (2010). The many shades of rose-coloured glasses: an evolutionary approach to the influence of different positive emotions. *Journal of Consumer Research*, 37(2):238 – 250.

Griskevicius, V., Tybur, J., Gangestad, S., Perea, E., Shapiro, J., and Kenrick, D. (2009). Aggress to impress: Hostility as an evolved context-dependent strategy. *Journal of Personality and Social Psychology*, 96(5):980 – 994.

Griskevicius, V., Tybur, J., Sundie, J., Cialdini, R., Miller, G., and Kenrick, D. (2007). Blatant benevolence and conspicuous consumption: When romantic motives elicit strategic costly signals. *Journal of Personality and Social Psychology*, 93(1):85 – 102.

Guilford, J. (1950). Creativity. *American Psychologist*, 5(9):444 – 454.

Guilford, J. (1967). *The nature of human intelligence*. McGraw-Hill, New York.

Hagen, E. and Bryant, G. (2003). Music and dance as a coalition signalling system. *Human Nature*, 14(1):21 – 51.

Hagen, E. and Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Musicae Scientiae Special issue on Music and Evolution*, 13(2):291 – 320.

Haldane, J. (1955). Animal communication and the origin of human language. *Science Progress*, 43(9-10):385 – 401.

Hare, B. and Tomasello, M. (2005). Human-like social skills in dogs? *Trends in Cognitive Sciences*, 9(9):439 – 444.

Hare, B., Wobber, V., and Wrangham, R. (2012). The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Animal Behaviour*, 83(3):573 – 585.

Hasson, O. (1994). Cheating signals. *Journal of Theoretical Biology*, 167(3):223 – 238.

Heaton, P. and Wallace, G. L. (2004). Annotation: the savant syndrome. *Journal of Child Psychology and Psychiatry*, 45(5):899 – 911.

Heim, A., Watts, K., and Simmonds, V. (1974a). *AH2/AH3 manual*. NFER Publishing, Windsor, UK.

Heim, A., Watts, K. P., and Simmonds, V. (1974b). AH2 and AH3: Parallel tests of reasoning. *British Journal of Psychology*, 65(4):493 – 503.

Henrich, J., Heine, S., and Norenzayan, A. (2010). The wierdest people in the world? *Behavioural and Brain Sciences*, 33(2 - 3):61 – 83.

Heyduk, R. (1975). Rated preference for musical compositions as it relates to complexity and exposure frequency. *Perception and Psychophysics*, 17(1):84 – 91.

Hill, A. L. (1977). Idiots savants: Rate of indicence. *Perceptual and Motor Skills*, 44(1):161 – 162.

Hill, R. and Dunbar, R. (2003). Social network size in humans. *Human Nature*, 14(1):53 – 72.

Hocevar, D. (1979). A comparison of statistical infrequency and subjective judgement as criteria in the meaurement of originality. *Journal of Personality Assessment*, 43(3):279 – 299.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203:88 – 96.

Hodgson, G. and Knudsen, T. (2006). The nature and units of social selection. *Journal of evolutionary economics*, 16(5):477 – 489.

Honda, E. and Okanoya, K. (1999). Acoustical and syntactical comparisons between songs of the White-backed Munia (Lonchura striata) and its domesticated strain, the Bengalese Finch (Lonchura striata var. domestica). *Zoological Science*, 16(2):319 – 326.

Hope, D., Bates, T., Gow, A., Starr, J., and Deary, I. (2010). Fluctuating asymmetry and personality. *Personality and Individual Differences*, 50(2011):49 – 52.

Hurford, J. (2003). The language mosaic and its evolution. In Christiansen, M. and Kirby, S., editors, *Language Evolution*, pages 38 – 57. Oxford University Press.

Huron, D. (2001). Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences*, 930:43 – 61.

Ingram, D. (1995). The cultural basis of prosodic modifications to infants and children: a response to Fernald's universalist theory. *Journal of Child Language*, 22(1):223 – 233.

Jablosnksi, N. (2004). The evolution of human skin and skin colour. *Annual Review of Anthropology*, 33:585 – 623.

Jaeggi, A., Burkart, J., and vanSchaik, C. (2010). On the psychology of cooperation in humans and other primates: combining the natural history and experimental evidence of prosociality. *Philosophical Transactions of the Royal Society of London B*, 365(1553):2723 – 2735.

Jarvis, E. (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science*, 5(1):5 – 14.

Jensen, A. (1998). *The g factor. The science of mental ability*. Praeger, Westport, Conneticut.

Jespersen, O. (1895). *Progress in Language: Amsterdam Classics on Linguistics 17*. John Benjamins, Amsterdam.

Jespersen, O. (1922). *Language: Its nature, development and origin*. Allen, London.

Kagawa, H., Yamada, H., Lin, R., Hasegawa, T., and Okanoya, K. (2010). Relaxation of selection can lead to signal variation: An example in birdsong. In Smith, A., Schouwstra, M., deBoer, B., and Smith, K., editors, *The evolution of language; Proceedings of the 8th international conference on the evolution of language*, pages 425 – 426. World Scientific, London.

Kemper, T. (1990). *Social structure and testosterone*. Rutgers University Press, New Brunswick, New Jersey.

Kiker, A. (2012). *Exploring the evolution of music and language: Social versus sexual selection. Evidence from an emotional priming study*. Unpublished MSc dissertation, University of Edinburgh.

Kimbel, W. (2009). The origin of Homo. In Grine, F., Fleagle, J., and Leakey, R., editors, *The first humans - origins and early evolution of the genus Homo*, pages 31 – 37. Springer, Berlin.

Kinsey, A., Pomeroy, W., and Martin, C. (1948). *Sexual behaviour in the Human Male*. W.B. Saunders, Philadelphia.

Kippen, J. and Bel, B. (1992). Modelling music with grammars: Formal language representation in the Bol Processor. In Marsden, A. and Pople, A., editors, *Computer Representations and Models in Music*, pages 207 – 238. Academic Press.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, 5(2):101 – 110.

Kirby, S. (2012). Darwin's musical protolanguage: An increasingly compelling picture. In Rebuschat, P., Rohrmeier, M., Hawkins, J., and Cross, I., editors, *Language and music as cognitive systems*, pages 96 – 102. Oxford University Press, Oxford.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31):10681 – 10686.

Kirby, S., Dowman, M., and Griffiths, T. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(2):5241 – 5425.

Kirby, S. and Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*. Springer, Berlin.

Kirby, S., Smith, K., and Brighton, H. (2004). From UG to universals: Linguistic adaptation through iterated learning. *Studies in Language*, 28(3):587 – 607.

Kirschner, S. and Tomasello, M. (2010). Joint music making promotes prosocial behaviour in 4 year old children. *Evolution and Human Behaviour*, 31(5):354 – 364.

Kleiman, D. (1977). Monogamy in mammals. *Quarterly Review of Biology*, 52(1):39 – 69.

Koelsch, S., Gunter, T. C., van Cramon, D. Y., Zysset, S., Lohmann, G., and Friederici, A. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage*, 17(2):956 – 966.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. (2004). Music, language, and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, 7(3):302 – 307.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2(2):205 – 254.

Lakoff, R. (1973). Language and woman's place. *Language in Society*, 2(1):45 – 79.

Leakey, R. (2009). Early humans: Of whom do we speak? In Grine, F., Fleagle, J., and Leakey, R., editors, *The first humans - origins and early evolution of the genus Homo*, pages 3 – 6. Springer, Berlin.

Lerdahl, F. and Jackendoff, R. (1983). *A generative theory of tonal music*. MIT Press, Massachusetts.

Levman, B. (1992). The genesis of music and language. *Ethnomusicology*, 36(2):147 – 170.

Lieberman, D. (2009). Brains, brawn and the evolution of human endurance running capabilities. In Grine, F., Fleagle, J., and Leakey, R., editors, *The first humans - origins and early evolution of the genus Homo*, pages 77 – 92. Springer, Berlin.

List, G. (1963). The boundaries of speech and song. *Ethnomusicology*, 7(1):1 – 16.

Livingstone, F. (1973). Did the Australopithecines sing? *Current Anthropology*, 14(1-2):25 – 26.

Longuet-Higgins, H. and Lee, C. (1984). Rhythm and meter: The rhythmic interpretation of monophonic music. *Music Perception*, 1(4):424 – 441.

Lowry, R. (2001). *Significance of the difference between two correlation coefficients*. Retrieved from http://vassarstats.net/rdiff.html, 16/8/2012.

MacLarnon, A. and Hewitt, G. (1999). The evolution of human speech: the role of enhanced breathing control. *American Journal of Physical Anthropology*, 109(3):341 – 363.

MacWhinney, B. (2002). The gradual emergence of language. In Givon, T. and Malle, B., editors, *The evolution of language out of pre-language*, pages 233 – 264. John Benjamins, Amsterdam.

Maess, B., Koelsch, S., Gunter, T., and Friederici, A. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, 4(5):540 – 545.

Malloch, S. M. (1999). Mothers and infants and communicative musicality. *Musicae Scientiae Special issue 1999-2000: Rhythm, Musical narrative and the Origins of Human Communication*, pages 29 – 57.

Manske, M. and Davis, G. (1968). Effects of simple instructional biases upon performance in the unusual uses test. *Journal of General Psychology*, 79(1):25 – 33.

Masataka, N. (2003). *The onset of language*. Cambridge University Press, Cambridge.

Masataka, N. (2009). The origins of language and the evolution of music: A comparative review. *Physics of life reviews*, 6(1):11 – 22.

Matychuk, P. (2005). The role of child-directed speech in language acquisition: a case study. *Language Sciences*, 27(3):301 – 379.

Mayer, R. (1999). Fifty years of creativity research. In Sternberg, R., editor, *Handbook of Creativity*, pages 449 – 460. Cambridge University Press, Cambridge.

Maynard Smith, J. and Harper, D. (1995). Animal signals: Models and terminology. *Journal of Theoretical Biology*, 177(3):305 – 311.

Maynard Smith, J. and Harper, D. (2003). *Animal signals*. Oxford University Press, Oxford.

Maynard Smith, J. and Parker, G. (1976). The logic of asymmetric contests. *Animal Behaviour*, 24(1):159 – 176.

McDermott, J. and Hauser, M. (2005). The origins of music: Innateness, uniqueness and evolution. *Music Perception*, 23(1):29–59.

McMullen, P. (1974). Influence of number of different pitches and melodic redundancy on preference responses. *Journal of Research in Music Education*, 22(3):198 – 204.

Merker, B. (2000). Synchronous chorusing and human origins. In N.L. Wallin, B. Merker, S. B., editor, *The origins of music*, pages 315 – 327. MIT Press, Massachusetts.

Merker, B., Madison, G., and Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, 45(1):4 – 17.

Mesoudi, A. and Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society of London B*, 363(1509):3489 – 3501.

Miller, G. (2000). Evolution of human music through sexual selection. In Wallin, N., Merker, B., and Brown, S., editors, *The origins of music*, pages 329 – 360. MIT Press, Massachusetts.

Miller, G. (2001). *The mating mind: How sexual choice shaped human nature*. Anchor Books, New York.

Miller, L. K. (1999). The savant syndrome: intellectual impairment and exceptional skill. *Psychological Bulletin*, 125(1):31 – 46.

Mithen, S. (2005). *The singing Neanderthals: The origins of music, language, mind and body*. Weidenfeld and Nicolson, London.

Mithen, S. (2012). The significance of stones and bones: understanding the biology and evolution of rhythm requires attention to the archaeological and fossil record. In Rebuschat, P., Rohrmeier, M., Hawkins, J., and Cross, I., editors, *Language and music as cognitive systems*, pages 103 – 109. Oxford University Press, Oxford.

Moller, A. P. and Swaddle, J. (1997). *Asymmetry, developmental stability and evolution*. Oxford University Press, Oxford.

Moore, J. (1981). *The Post-Darwinian Controversies: A Study of the Protestant Struggle to Come to Terms with Darwin in Great Britain and America, 1870-1900*. Cambridge University Press, Cambridge.

Nakata, T. and Trehub, S. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behaviour and development*, 27(4):455 – 464.

Nelson, D. A. (2000). A preference for own-subspecies' song guides vocal learning in a song bird. *Proceedings of the National Academy of Sciences USA*, 97(24):13348 – 13353.

Nesse, R. (2007). Runaway social selection for displays of partner value and altruism. *Biological Theory*, 2(2):143–155.

Nesse, R. (2009). Social selection and the origins of culture. In Schaller, M., Heine, S., Norenzayan, A., Yamagishi, T., and Kameda, T., editors, *Evolution, culture, and the human mind*, pages 137 – 150. Lawrence Erblaum Associates, Philadelphia.

Nettl, B. (1969). *Music in primitive culture*. Harvard University Press, Cambridge, MA.

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In Wallin, N., Merker, B., and Brown, S., editors, *The origin of music*, pages 463 – 472. MIT Press, Massachusetts.

Nettle, D. (2009). *Evolution and genetics for psychology*. Oxford University Press, Oxford.

Nettle, D. and Dunbar, R. I. M. (1997). Social markers and the evolution of reciprocal exchange. *Current Anthropology*, 38(1):93 – 98.

Neuhaus, C., Knosche, T., Bahlman, J., and Friederici, A. (2009). Grammar types in language explain tone sequence processing in music. In Louhivouri, J., Eerola, T., Saarikallio, S., Himberg, T., and Eorbla, P., editors, *Proceedings of the 7th Triennial conference of European Society for the Cognitive Sciences of Music*, pages 372 – 377. ESCOM 2009.

Newport, E., Gleitman, H., and Gleitman, L. (1977). Mother, I'd rather do it myself: some effects and non-effects of maternal speech style. In Snow, C. and Ferguson, C., editors, *Talking to children*, pages 109 – 149. Cambridge University press.

North, A. and Hargreaves, D. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology*, 14(1-2):77 – 93.

Nowak, M. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805):1560 – 1563.

Nowak, M. and Krakauer, D. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028 – 8033.

Nowak, M. and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063):1291 – 1298.

Nowicki, S., Searcy, W., and Peters, S. (2002). Brain development, song learning and mate choice in birds: a review and experimental test of the “nutritional stress hypothesis”. *Journal of Comparative Physiology A*, 188(11-12):1003 – 1014.

Nwokah, E. (1987). Maidese versus Motherese - is the language input of child and adult carers similar? *Language and Speech*, 30(3):213 – 237.

Ochs, E. (1988). *Culture and language development: Langage acquisition and language socialisation in a Samoan village*. Cambridge University Press, Cambridge.

Okanoya, K. (2002). Sexual display as a syntactic vehicle. In Wray, A., editor, *The Transition to Language*, pages 46 – 63. Oxford University Press, Oxford.

Okanoya, K. (2004). The Bengalese finch: A window on the behavioural neurobiology of birdsong syntax. *Annals of the New York Academy of Sciences*, 1016(1):724 – 735.

Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1):97 – 113.

Opacic, T., Stevens, C., and Tillmann, B. (2009). Unspoken knowledge: Implicit learning of structured human dance movement. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 35(6):1570 – 1577.

Palmer, A. R. (1996). Waltzing with asymmetry. *Bioscience*, 46(7):518 – 532.

Patel, A. (2008). *Music, language, and the brain*. Oxford University Press, Oxford.

Patel, A. and Iversen, J. (2003). Acoustic and perceptual comparison of speech and drum sounds in the North Indian Tabla tradition: An empirical study of sound symbolism. In *Proceedings of the 15th International Congress of Phonetic Sciences, 3-9 August 2003, Bercelona*, pages 925 – 928, Barcelona.

Patel, A. and Iversen, J. (2006). A non-human animal can drum a steady beat on a musical instrument. In Baroni, M., Addessi, A., Caterina, R., and Costa, M., editors, *Proceedings of the 9th International Conference on Music Perception and Cognition*, page 477, Bologna. University of Bologna.

Patel, A., Iversen, J., Bergman, M., and Schultz, I. (2009). Experimental evidence for synchronisation to a musical beat in a non-human animal. *Current Biology*, 19(10):827 – 830.

Patel, A., Iversen, J., Bregman, M., Schulz, I., and Schulz, C. (2008). Investigating the human-specificity of synchronization to music. In Adachi, M., Nakajima, Y., and Hiraga, Y., editors, *Proceedings of the 10th International Conference on Music Perception and Cognition*, pages 100 – 104, Adelaide. Causal Productions.

Peek, P. and Yankah, K. (2004). *African folklore: An encyclopaedia*. Taylor and Francis, London.

Penke, L. and Asendorpf, J. (2008). Beyond global sociosexual orientations: A more differential look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, 95(5):1113 – 1135.

Penke, L., Bates, T., Gow, A., Pattie, A., Starr, J., Jones, B., Perrett, D., and Deary, I. (2009). Symmetric faces are a sign of successful cognitive aging. *Evolution and human behaviour*, 30(6):429 – 437.

Penke, L., Maniega, S. M., Bastin, M., Valdés Hernández, M., Murray, C., Royle, N., Starr, J., Wardlaw, J., and Deary, I. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry*, 17(10):1026 – 1030.

Pinker, S. (1997). *How the mind works*. W.W. Norton, New York.

Plantinga, J. and Trainor, L. (2009). Melody recognition by two-month-old infants. *Journal of the Acoustical Society of America*, 125(2):EL58 – EL62.

Poole, J. H., Tyack, P., Stoeger-Horwath, A., and Watwood, S. (2005). Animal behaviour: Elephants are capable of vocal learning. *Nature*, 434(7032):455 – 456.

Pothos, E. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, 133(2):227 – 244.

Prokosch, M., Yeo, R., and Miller, G. (2005). Intelligence tests with higher g loadings show higher correlations with body symmetry: evidence for a general fitness factor mediated by developmental stability. *Intelligence*, 33(2):203 – 213.

Ralls, K. (1977). Sexual dimorphism in mammals: Avian models and unanswered questions. *American Naturalist*, 111(981):917 – 938.

Ramscar, M. and Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science*, 31(6):927 – 960.

Ratner, N. B. and Pye, C. (1984). Higher pitch in BT is not universal: acoustic evidence from Quiche Mayan. *Journal of Child Language*, 11(3):512 – 522.

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour*, 6(6):855 – 863.

Reiss, N. (1989). Speech act taxonomy, chimpanzee communication, and the evolutionary basis of language. In Wind, J., Pulleybank, E., DeGrolier, E., and Bichakjian, B., editors, *Studies in Language Origins*, pages 283 – 304. John Benjamins, Amsterdam.

Rimland, B. (1978). Savant capabilities of autistic children and their cognitive abilities. In Serban, G., editor, *Cognitive defects in the development of mental illness*, pages 41 – 64. Brunner-Mazel, New York.

Rimm, S. (1976). Gift: An instrument for the identification and measurement of creativity. *Dissertation Abstracts International*, 37(5-A):2804.

Rimm, S. and Davis, G. (1976). Gift: An instrument for the identification of creativity. *Journal of Creative Behaviour*, 10(3):178 – 182.

Ritchie, G. (2009). *Modelling the transition to complex, culturally transmitted communication*. PhD thesis, University of Edinburgh.

Ritchie, G. and Kirby, S. (2005). Selection, domestication and the emergence of learned communication systems. In *2nd international symposium on the emergence and evolution of linguistic communication*, University of Hertfordshire, Hatfield. AISB.

Ritchie, G. and Kirby, S. (2006). Modelling the transition to learned communication: An initial investigation into ecological conditions favouring cultural transmission. In Cangelosi, A., Smith, D., and Smith, K., editors, *The Evolution of Language: Proceedings of the 6th international conference on the evolution of language*, pages 283 – 291, London/ New York:. World Scientific.

Ritchie, G. and Kirby, S. (2007). A possible role for selective masking in the evolution of complex learned communication systems. In Lyon, C., Nehaniv, C., and Cangelosi, A., editors, *Emergence of communication and language*, pages 387 – 401. Springer Verlag, London.

Ritchie, G., Kirby, S., and Hawkey, D. (2008). Birdsong as an indicator mechanism: Modelling the developmental stress hypothesis. *Journal of Theoretical Biology*, 251(4):570 – 583.

Roberts, G. (2010). An experimental study of social selection and frequency of interaction in linguistic diversity. *Interaction Studies*, 11(1):138 – 159.

Roche, H., Blumenschine, R., and Shea, J. (2009). Origins and adaptations of early Homo: what archaeology tells us. In Grine, F., Fleagle, J., and Leakey, R., editors, *The first humans - origins and early evolution of the genus Homo*, pages 135 – 147. Springer, Berlin.

Russett, C. (1976). *Darwin in America: the intellectual response 1865 - 1912*. W.H. Freeman and Co. Ltd., San Francisco.

Schachner, A., Brady, T., Pepperberg, I., and Hauser, M. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, 19(10):831 – 836.

Schaefer, C. (1969). The prediction of creative achievement from a biographical inventory. *Educational and Psychological Measurement*, 29(2):431 – 437.

Schaller, G. (1963). *The mountain gorilla*. University of Chicago Press, Chicago.

Schenker, H. (1954). *Harmony*. E. Mann-Borgese (transl.), O.Jones (ed.). Chicago University Press.

Schenker, H. (1979). *Free composition (Der Freie Satz)*. E. Ostler (transl.), O. Jones (ed.). Longman, New York.

Schieffelin, B. (1985). *The acquisition of Kaluli*. The crosslinguistic study of language acquisition. Erblaum.

Schmulevich, I. and Povel, D. (2000). Measures of temporal pattern complexity. *Journal of New Music Research*, 29(1):61 – 69.

Schoenberg, A. (1975). *Style and Idea: Selected writings of Arnold Schoenberg*. St. Martin's Press, New York.

Seyfarth, R., Cheney, D., and Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, 210(4471):801 – 803.

Shute, B. and Wheldall, K. (1989). Pitch alterations in British motherese: some preliminary acoustic data. *Journal of Child Language*, 16(3):503 – 512.

Shute, B. and Wheldall, K. (1995). The incidence of raised average pitch and increased pitch variability of British motherese speech and the influence of maternal characteristics and discourse form. *First Language*, 15(1):35 – 55.

Simmons, L., Rhodes, G., Peters, M., and Koehler, N. (2004). Are human preferences for facial symmetry focused in signals of developmental instability? *Behavioural Ecology*, 15(5):864 – 871.

Simner, J., Cuskley, C., and Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception*, 39(4):553 – 569.

Simonton, D. (1980a). Thematic fame and melodic originality in classical music: A multivariate computer-content analysis. *Journal of Personality*, 48(2):206 – 219.

Simonton, D. (1980b). Thematic fame, melodic originality and musical zeitgeist: A biographical and transhistorical content analysis. *Journal of Personality and Social Psychology*, 38(6):972 – 983.

Skoyles, J. (2000a). The singing origin theory of speech. In *The evolution of language; Proceedings of the 3rd international conference on the evolution of language: EVOLANG 3*, Paris.

Skoyles, J. (2000b). Without breath and without song? *Trends in Cognitive Sciences*, 4(11):405.

Slater, P. (2000). Birdsong repertoires: their origins and use. In Wallin, N., Merker, B., and Brown, S., editors, *The origins of music*, pages 49 – 64. MIT Press, Massachusetts.

Smith, J. and Melara, R. (1990). Aesthetic preference and syntactic prototypicality in music: 'tis the gift to be simple. *Cognition*, 34(3):279 – 298.

Smith, K., Kirby, S., and Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, 9(4):371 – 386.

Sober, E. and Wilson, D. (1998). *Unto others: the evolution and psychology of unselfish behaviour*. Harvard University Press, Cambridge, MA.

Speller, K. and Schumacher, G. (1975). Age and set in creative test performance. *Psychological Reports*, 36(2):447 – 450.

Spencer, K., Buchanan, K., Goldsmith, A., and Catchpole, C. (2003). Song as an honest signal of developmental stress in the zebra finch (Taeniopygia gutta). *Hormones and Behaviour*, 44(2):132 – 139.

Sponheimer, M. and Lee-Thorp, J. (2009). Biochemical evidence for the environments of early Homo in South Africa. In Grine, F., Fleagle, J., and Leakey, R., editors, *The first humans - origins and early evolution of the genus Homo*, pages 185 – 194. Springer, Berlin.

Steck, L. and Machotka, P. (1975). Preference for musical complexity: Effects of context. *Journal of Experimental Psychology: Human perception and performance*, 104(2):170 – 174.

Stein, M. (1953). Creativity and culture. *Journal of Psychology*, 36(2):311 – 322.

Tamariz, M., Brown, J., and Murray, K. (2010). The role of practice and literacy in the evolution of linguistic structure. In Smith, A., Schouwstra, M., deBoer, B., and Smith, K., editors, *The evolution of language; Proceedings of the 8th international conference on the evolution of language: EVOLANG 8*, pages 313 – 320, London. World Scientific.

Tanaka, Y. (1996). Social selection and the evolution of animal signals. *Evolution*, 50(2):512 – 523.

Tannen, D. (1990). *You just don't understand: Women and men in conversation*. Virago Press, London.

Taylor, D. (1963). Variables related to creativity and productivity among men in two research laboratories. In Taylor, C. and Barron, F., editors, *Scientific creativity: Its recognition and development; Selected papers from the proceedings of the first, second, and third University of Utah conferences: "The Identification of Creative Scientific Talent"*. Wiley, New York.

Thompson, B., Smith, K., and Kirby, S. (2012). Cultural evolution renders linguistic nativism implausible. In Scott-Phillips, T., Tamariz, M., Cartmill, E., and Hurford, J., editors, *The* *evolution of language; Proceedings of the 9th international conference on the evolution of language*, pages 557 – 558, London. World Scientific.

Thorndike, E. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1):25 – 29.

Torrance, E. (1962). *Guiding creative talent*. Prentice Hall, Englewood Cliffs. N.J.

Trainor, L. (2007). Do preferred beat rate and entrainment to the beat have a common origin in movement? *Empirical Musicology Review*, 2(1):17 – 20.

Trainor, L. and Unrau, A. (2012). Development of pitch and music perception. In Werner, L. and Popper, R. F. . A., editors, *Springer Handbook of Auditory Research: Human auditory development*, pages 223 – 254. Springer Science.

Treffert, D. A. (2009). The savant syndrome: an extraordinary condition. a synopsis: past, present, future. *Philosophical Transactions of the Royal Society of London B*, 364(1522):1351 – 1357.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae Special issue 1999-2000: Rhythm, Musical narrative and the Origins of Human Communication*, pages 155–215.

Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46(1):35 – 57.

Trivers, R., Palestis, B., and Zaatari, D. (2009). *The anatomy of a fraud: Symmetry and dance*. TPZ Publishers, Antioch, California.

Trut, L. (1999). Early canid domestication: The farm-fox experiment. *American Scientist*, 87(2):160 – 169.

Trut, L. (2001). Experimental studies of early canid domestication. In Ruvinsky, A. and Sampson, J., editors, *The genetics of the dog*, pages 15 – 42. CABI Publishing.

Ungar, P. (2004). Dental topography and diets of Australopithecus afarensis and early Homo. *Journal of Human Evolution*, 46(5):605 – 622.

Ungar, P. and Scott, R. (2009). Dental evidence for diets of early Homo. In Grine, F., Fleagle, J., and Leakey, R., editors, *The first humans - origins and early evolution of the genus Homo*, pages 121 – 134. Springer.

VanDongen, S. and Gangestad, S. (2011). Human fluctuating asymmetry in relation to health and quality: a meta-analysis. *Evolution and Human Behaviour*, 32(6):380 – 398.

Verhoef, T., deBoer, B., and Kirby, S. (2012). Holistic or synthetic protolanguage: Evidence from iterated learning of whistled signals. In Scott-Phillips, T., Tamariz, M., Cartmill, E., and Hurford, J., editors, *The Evolution of Language: Proceedings of the 9th international conference on the evolution of language*, pages 368 – 375, London. World Scientific.

Verhoef, T., Kirby, S., and Padden, C. (2011). Cultural emergence of combinatorial structure in an artificial whistled language. In Carlson, L., Holscher, C., and Shipley, T., editors, *Proceedings of Cognitive Science Society Conference 33*, pages 483 – 488, Austin, Texas. Cognitive Science Society.

Vitz, P. (1964). Preference for rates of information presented by sequences of tones. *Journal of Experimental Psychology*, 68(2):176 – 183.

Volkmar, F. R. and Lord, C. (1999). Diagnosis and definition of autism and other pervasive developmental disorders. In Volkmar, F., editor, *Autism and pervasive developmental disorders*, pages 1 – 31. Cambridge University press, Cambridge.

Wallen, K. and Lovejoy, J. (1993). Sexual behavior. In Schulkin, J., editor, *Hormonally Induced Changes in Mind and Brain*, pages 71 – 97. Academic Press, San Diego.

Wallin, N., Merker, B., and Brown, S., editors (2000). *The Origins of Music*. MIT Press.

Watson, P. and Thornhill, R. (1994). Fluctuating asymmetry and sexual selection. *Trends in ecology and evolution*, 9(1):21 – 25.

Webb, G. (1994). *The evolution controversy in America*. University of Kentucky, Kentucky.

Weiss, G. (1974). On Livingstone's 'Did the Australopithecines sing?'. *Current Anthropology*, 15(1-2):103 – 104.

Welch, G. F. (1998). Early childhood musical development. *Research Studies in Music Education*, 11(1):27 – 41.

West-Eberhard, M. (1979). Sexual selection, social competition, and evolution. *Proceedings of the American Philosophical Society*, 123(4):222 – 234.

West-Eberhard, M. (1983). Sexual selection, social comoetition and speciation. *Quarterly Review of Biology*, 58(2):155 – 183.

Windsor, L. and Desain, P. (2000). Introduction: Multiple perspective on rhythm perception and production. In Desain, P. and Windsor, L., editors, *Rhythm: Perception and production*, pages xi – xvi. Swets and Zeitlinger, Lisse, The Netherlands.

Wolf, J., Brodie, E., and Moore, A. (1999). Interacting phenotypes and the evolutionary process II. Selection resulting from social interactions. *American Naturalist*, 153(3):254 – 266.

Wood, B. (2009). Where does the genus Homo begin, and how would we know? In Grine, F., Fleagle, J., and Leakey, R., editors, *The first humans - origins and early evolution of the genus Homo*, Vertebrate Paleobiology and Paleoanthropology, pages 17 – 28. Springer, Berlin.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, 18(1):47 – 67.

Wray, A. (2000). Holistic utterances in protolanguage: the link from primates to humans. In Knight, C., Studdert-Kennedy, M., and Hurford, J., editors, *The evolutionary emergence of language: Social function and the origins of linguistic form*, pages 285 – 302. Cambridge University Press, Cambridge.

Wray, A. (2002a). Dual processing in protolanguage: Performance without competence. In Wray, A., editor, *The transition to language*, pages 113 – 137. Oxford University Press, Oxford.

Wray, A. (2002b). *Formulaic language and the lexicon*. Cambridge University Press, Cambridge.

Wray, A. (2002c). *The transition to language*. Oxford University Press.

Wynn, T. (1998). Did Homo erectus speak? *Cambridge Archaeological Journal*, 8(1):78 – 81.

Yeo, R., Hill, D., Campbell, R., Vifil, J., and Brooks, W. (2000). Developmental instability and working memory ability in children: A magnetic resonance spectroscopy investigation. *Developmental Neuropsychology*, 17(2):143 – 159.

Zahavi, A. (1975). Mate selection - a selection for a handicap. *Journal of Theoretical Biology*, 53(1):205 – 214.

Zahavi, A. and Zahavi, A. (1997). *The handicap principle*. Oxford University Press, Oxford.

Ziv, N. and Keydar, E. (2009). The relationship between creative potential, aesthetic response to music, and musical preferences. *Creativity Research Journal*, 21(1):125 – 133.