Modelling the transition to complex, culturally transmitted communication

Graham R.S. Ritchie

B.Sc.(Hons), M.Sc.

Doctor of Philosophy

School of Informatics

University of Edinburgh

2009

Abstract

Human language is undoubtedly one of the most complex and powerful communication systems to have evolved on Earth. Study of the evolution of this behaviour is made difficult by the lack of comparable communication systems elsewhere in the animal kingdom, and by the fact that language leaves little trace in the fossil record. The human language faculty can, however, be decomposed into several component abilities and a proposed evolutionary explanation of the whole must address (at least) the evolution of each of these components. Some of these features may also be found in other species, and thus permit use of the powerful comparative method. This thesis addresses the evolution of two such component features of human language; complex vocal signalling and the cultural transmission of these vocal signals.

I argue that these features make a significant contribution to the nature of human language as we observe it today and so a better understanding of the evolutionary processes that gave rise to them will contribute to study of the evolution of language. This thesis addresses the evolution of these features firstly by identifying other communication systems found in nature that display them, and focusing in particular on the song of the oscine passerines (songbirds). Bird song is chosen as a model system because of the wealth of empirical data on nearly all aspects of the behaviour and the variety of song behaviour found in this group. There also appear to be some striking similarities in the development of language and song. I argue that a better understanding of the evolution of complex signalling and cultural transmission in songbirds and other species will provide useful insight into the evolution of these features in language. This thesis presents a series of related formal models that investigate several issues in the evolution of these features.

I firstly present a simple formal model of bird song acquisition and use this in a computational model of evolution to investigate some ecological conditions under which vocal behaviour can become more or less reliant on cultural transmission. I then present a pertinent case study of two closely related songbird sub-species and develop a computational model that demonstrates that domestication, or a similar shift in the fitness landscape, may play a surprising role in the evolution of signal complexity (in some sense) and increased vocal plasticity. Finally, I present several models that investigate the plausibility and consistency of the ‘developmental stress hypothesis’, an important hypothesis drawn from the biological literature that proposes that song learning and song complexity may serve as a sexually selected mate quality indicator mechanism. These models provide the first theoretical support for this important but complex hypothesis and identify a number of relevant parameters that may affect the evolution of such a system.

Acknowledgements

Many thanks to my supervisor, Simon Kirby, for his ideas, guidance and support throughout the development of this thesis. Thanks also to others who have given me guidance over the course of my PhD studies, especially John Levine, Jim Hurford and Jon Oberlander.

The LEC was a great place to work as postgraduate student. Thanks to Dan Dediu, Anna Kinsella, Dave Hawkey, Stefan Hoefler, Thom Scott-Phillips, Jelle Zuidema, Kenny Smith, Andrew Smith, Henry Brighton, Hajime Yamauchi, Monica Tamariz, Paul Vogt and all the other members of the LEC for invaluable discussions, comments and support, and for providing such a stimulating research environment. Thanks also to Richard Carter for the motivational Tuesday lunches, and for useful comments on some of the work presented here.

I spent a very enjoyable month at the RIKEN Brain Science Institute in Kazuo Okanoya's Biolinguistics lab. Thanks to Prof. Okanoya and everyone else there for their hospitality and for several very useful discussions, and to the Japan Society for the Promotion of Science for providing funding. Thanks also to the Santa Fe Institute for funding and hosting a stimulating Complex Systems Summer School. Terrence Deacon provided some insightful suggestions and comments on the model presented in chapter 4 of this thesis for which I am also very grateful.

Last but certainly not least, my love and thanks to Lottie, for all her support and for sticking it out over these past few years.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Graham R.S. Ritchie*
B.Sc.(Hons), M.Sc.)

Table of Contents

5 Song learning as an indicator mechanism 77

6 Conclusions 117

Chapter 1

Introduction & motivation

1.1 What's so special about language

Language is one of the most distinctive characteristics of our species. Much modern work in the biological and psychological sciences has identified profound similarities in almost all living species, from the universality of the genetic code to similarities in morphogenesis, development and behaviour. However, as yet we have found no other species that has a communication system that is as complex or as powerful as human language. Indeed, Maynard Smith and Szathmáry (1995) recently characterised the evolution of language, alone among the myriad communication systems found in nature, as one of the major transitions in evolution of life, alongside other fundamental innovations such as the evolution of the genetic code and the split of the eukaryotes from the prokaryotes. They give language this prominent position because, they argue, it provides a new mechanism for the transmission of information.

One feature of human language that marks it out from the communication of many other species (though not all) is the extent to which it relies on post-natal experience. There is undoubtedly a significant genetic contribution to human language abilities, but a child must be exposed to adult language in the first few years of life if he is to attain a normal language ability. Adult language behaviour is clearly strongly affected by this early exposure; a child raised in Japan will learn to produce very different sounds to a child raised in the UK. This is in marked contrast to the vocal communication of our closest extant relatives the chimpanzees which have stereotypical calls with very little apparent geographic variation (reviewed in Janik and Slater, 1997). It therefore appears that in the time since the last common ancestor of humans and chimpanzees we have acquired the ability to modify our vocalisations as a result of experience, an ability known to biologists as *vocal learning* (Janik and Slater, 1997). Our vocal learning abilities provide us with a uniquely flexible communication system which can adapt as our communicative requirements change.

As Boyd and Richerson (1985) note, nearly all organisms appear to be able to vary their phenotypes adaptively in response to environmental conditions. What makes language learning different from this ubiquitous phenotypic plasticity is that language is acquired from other individuals and is transmitted from generation to generation via means other than genes. This extra-genetic, or cultural, transmission of language means that variants acquired by individual learning or other forms of phenotypic plasticity are not lost when the individual dies but, like genes, can be transmitted to the next generation (Boyd and Richerson, 1985). The features of language that are transmitted in this way may then evolve in a manner analogous to, but separate from, genetic evolution. A number of authors have argued that the cultural evolution of language may provide an alternative source of explanation to biological evolution for some important features of language (e.g. Deacon, 1997; Kirby, 1999), and I will discuss some of these proposals in chapter 2.

The structure of language as we observe it thus results from the interactions of at least three complex systems; biological genetic evolution, individual ontogenetic development, and cultural evolution. The interaction of these three systems is depicted in figure 1.1. One of Deacon's (1997) central arguments is that language structure is the result of a complex co-evolutionary process in which the human brain has evolved to fit the unique requirements of language learning and use, and language has adapted by cultural evolution to fit the biases and constraints of the brains that acquire it. If Deacon is right and such a co-evolutionary process makes an important contribution to the structure of language, an important question for the study of language evolution is how this co-evolutionary process got started. A system of cultural evolution in language requires that humans are capable of learning features of language from each other. As

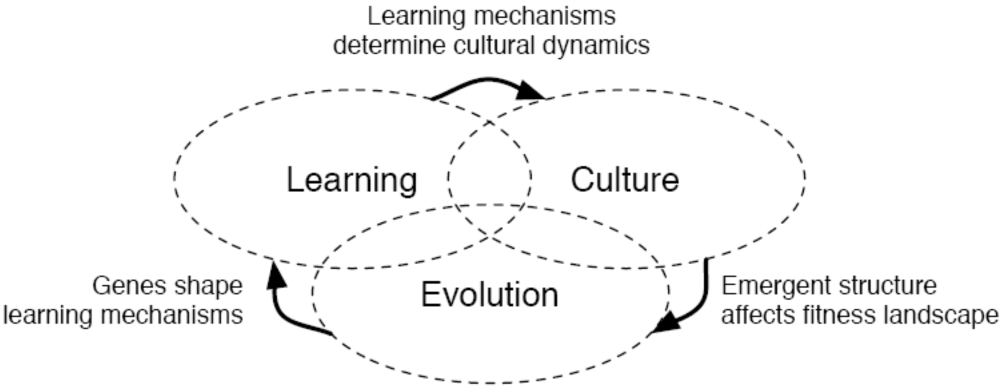


Figure 1.1: A graphical representation of the three complex dynamic systems that contribute to language evolution; biological genetic evolution, individual development and cultural transmission. Figure taken from (Kirby, 2007).

noted above, this is an ability that is not shared with many other species including our closest relatives. The origin of vocal learning in humans was therefore a significant transition in the evolution of language and this is the central topic of this thesis.

1.2 Guide to the thesis

Two key abilities underly the cultural transmission of language in humans; namely vocal learning and complex vocal signalling. Vocal learning allows us to acquire features of language from other individuals and a capacity for using complex vocal signals gives language its expressive power. This thesis investigates the evolution of these features in communication systems with a series of related formal models.

Chapter 2 describes these two features in detail, and demonstrates their important contribution to several phenomena observed in human language. This chapter also reviews the occurrence and some proposed functions of these features in other species and contrasts these with human vocal behaviour. The aim of such comparisons is to identify any common themes that will inform our hypotheses about the evolution and function of these features in humans. Of other animal communication systems that display these two features, bird song is identified as an intriguing model system because the song of songbirds is also culturally transmitted by iterated learning and there are some striking parallels between language and song. These parallels suggest that there may be strong constraints on the evolution of a system of complex vocal learning which implies that a better understanding of the evolution of these two features in bird song will inform our understanding of their origin in humans.

This thesis then investigates a number of issues in the evolution of the cultural transmission of complex signals in bird song with a series of related formal models. **Chapter 3** introduces a simple formal framework for modelling the evolution of vocal learning in songbirds. I present a simple computational implementation of the most widely accepted model of song acquisition, the auditory template model. I then use this implementation in a computational agent-based model to investigate some ecological factors that are likely to play a role in the evolution of vocal learning abilities. In particular the models demonstrate that vocal learning is only likely to evolve if cultural transmission is sufficiently reliable, and when an animal does not require communication abilities until adulthood.

Chapter 4 presents a relevant case study of the evolution of vocal complexity and increased influence from cultural transmission in a domesticated songbird species, the Bengalese finch. Surprisingly, the feral progenitor of this species sings a markedly simpler and more stereotypical song. I present a computational model inspired by these data that demonstrates how increased song complexity and influence from learning could arise without direct selection for such abilities. The model demonstrates these behaviours may arise due to the *masking* of selection pressure, in this case due to domestication. The results suggest that a similar major change in the fitness landscape of early humans may explain how our communicative abilities may have diverged from the largely genetically specified systems of our close relatives.

Chapter 5 investigates a recent novel hypothesis for the evolutionary function of song learning, known as the ‘developmental stress hypothesis’. This hypothesis proposes that an individual’s ability to acquire complex vocal signals may serve as a sexually selected indicator mechanism, providing information to potential mates about an individual’s early development. I set this hypothesis in the context of other work on sexually selected signals. I then present the first theoretical investigation of this hypothesis, in the form of both computational and more general mathematical models.

The models demonstrate that, under some conditions, such a system can evolve and will be evolutionarily stable. I also discuss some wider implications of this hypothesis. Finally **chapter 6** concludes the thesis, identifying its main contributions and discussing outstanding issues in the evolution of human speech and language.

Chapter 2

Language in the context of other animal communication systems

In this chapter I set human language in the context of other communication systems found in nature, identifying some important features of human language that are shared with other species. I look in particular at two key features of human speech; *cultural transmission* and *complex vocalisation*, and their occurrence in other species, particularly songbirds, as these features will be the focus of this thesis. The aim of such comparisons is to demonstrate that insights and arguments proposed for the evolution of such behaviours in other species may be fruitfully applied to questions on the evolution of human linguistic behaviour.

2.1 Decomposing the language faculty

Communication systems abound in the natural world (e.g. Hauser, 1996), but the human language faculty does appear to be unique in the sense that no other species (as far as we currently know) uses or has used culturally transmitted syntactic vocal signals to convey arbitrary complex semantic information. This makes language a fascinating topic of study, but poses some problems for the researcher interested in the evolutionary origins of language, as one standard approach to tackling questions about the evolution of some feature is to look for other species that share this feature. This *comparative method* allows the researcher to pose and test hypotheses about the adaptive significance of a feature using "natural evolutionary experiments" provided by convergent evolution (Futuyma, 2005). Where a similar feature has evolved independently in several lineages, it is often possible to infer the function of the feature by identifying common selective factors, such as ecology, which are shared between the different lineages. For example the independent evolution of large body size in northern populations of many mammals suggests that large body size is an adaptation to colder climates (Futuyma, 2005). Identifying convergently evolved features may also allow glimpses into the possible evolutionary trajectory of the feature as the different lineages may have developed the feature to a greater or lesser extent. The comparative method is one of the most powerful methods in evolutionary biology as it can provide insights at several different levels of enquiry. Another powerful approach to studying the evolution of some feature is to look for evidence of intermediate forms in the fossil record. Unfortunately for the evolutionary linguist, utterances and the brains that produce and process them leave little for archaeologists to find, though changes in human physiology supporting speech are amenable to this type of study.

These observations would appear to deny a standard empirical approach to studying the evolution of language. However, as many authors have noted (e.g. Hockett, 1960; Hauser et al., 2002; Okanoya, 2002; Hurford, 2003; Hauser et al., 2002; Fitch, 2005), language is a complex behaviour which relies on a number of component abilities. Each of these components may have somewhat separate evolutionary histories, and crucially, some of these components may be shared to some extent with other species and therefore allow us to use the power of the comparative method.

Different authors disagree on the nature and uniqueness of these components, but many researchers (e.g. Fitch, 2000; Hauser, 2001) agree that the most common medium of language, speech, may be studied independently of other linguistic behaviours. The existence of other media for language, such as sign language and writing, demonstrate that speech does appear to be somewhat separable from other components supporting language. Human speech is still a complex behaviour that differs substantially in some ways from the vocal behaviour of our closest ancestors, the non-human primates, but a large proportion of the physiology underlying human speech is fundamentally similar to other animal vocal systems (Fitch, 2000).

Two features of speech in particular appear to mark human vocal behaviour out from that of other, even closely related, species:

- The complexity of our vocal signals
- The acquisition of these complex vocal signals by cultural transmission

It is these two features of human speech that concern this thesis, and the following sections explore them in more detail. It turns out that while these two features are comparatively rare in non-human animal communication systems, they are both found to some extent in several other species and I will review the biological literature covering their occurrence.

These two features are not the only interesting or important components of human speech, or the human language faculty as a whole, and an understanding of the evolution of these two traits will only form part of a full understanding of the evolution of language. In particular the way in which humans are able to use their culturally acquired vocal signals to communicate arbitrary and complex propositional content to one another is an ability that appears to be unique to our species, and one that we seem to be some way from providing a full evolutionary explanation for. However, as I argue below, these two traits make a significant contribution to the nature of language as we observe it and have some far-reaching consequences for human communication.

2.2 Complex vocalisation

2.2.1 Human vocal complexity

In some sense we can characterise all vocalisations as complex; vocalisations are complicated airborne (or sometimes waterborne) waveforms which have resulted from a series of muscular motor actions, which generally act to force air through specialised organs which generate and modify this waveform as the air passes through. These motor actions are initiated by neural processes that we are only just beginning to understand. This is not to mention the evolutionary processes which gave rise to the ability to produce such behaviour. While this is clearly as true for humans as it is for many other species, speech does have some features which mark it out from many other animal vocalisations and other human vocalisations such as crying and laughter.

Speech distinguishes itself from other vocalisations with the hierarchical and structured way in which acoustically distinct units of sound are combined together (Jackendoff, 2002). Meaningless phonemes are combined in regular patterns into meaningful words (or more correctly, morphemes). Morphemes are then themselves combined into entire sentences or utterances. Phonemes are not randomly distributed in human speech, but adhere to the phonotactic rules for the language being spoken. Words are also clearly not combined at random, and exploration of the subtleties of the many apparent rules guiding the combination of morphemes ('syntax') constitutes a large proportion of work in linguistics. This *duality of patterning* was one of four features Hockett (1960) identified as unique to human language. It is this hierarchical and combinatorial nature of speech that I wish to categorise as 'complex'.

Combinatorial structure plays a role in at least two ways in human speech, but what are the advantages of organising a communication system in this way? Among other functions, humans use speech to convey information about the world to each other and so we need some way to map meanings to (speech) signals. A simple system would be to define a repertoire of signals and pair these off with each required meaning. Such a system may work for a limited set of required meanings, and this may be a reasonable model of the alarm call behaviour of vervet monkeys (*Chlorocebus pygerythrus*) (Cheney and Seyfarth, 1990) which appear to have three distinct alarm calls (signals) which they use to warn conspecifics about three classes of predator (meanings). Humans, however, are able to communicate an apparently unlimited set of meanings. For us such a system would quickly require an unmanageably large repertoire of distinct signals. Speech (as well as alphabetic writing systems) resolves this problem by using a set of small signalling units that can be reused and recombined to produce a (theoretically) unlimited set of signals, a feature which Hockett (1960) terms 'productivity'. A signalling system that uses a small set of distinct discrete units may also support more robust signals as a small error in the production of one unit is unlikely to result in another valid but incorrect unit, which would be likely with a more ‘crowded’ signal space (de Boer, 2005). Such a combinatorial system may also make more storage of a large repertoire of signals more efficient, as instead of storing the complete detail of each and every signal, only the signalling units must be stored in detail and full signals can be stored as strings over these units. de Boer (2005) shows with a computational model that in an environment where it is advantageous to maintain a set of signals using as little storage as possible, a population of agents is indeed more likely to evolve a combinatorial phonemic system than a holistic system where the details of every signal need to be stored.

It seems then that a significant pressure for vocal complexity in human speech stems from the requirement that speech be sufficiently expressive to convey the range of messages we wish to share with each other. Other pressures may also have played a role though.

2.2.2 Vocal complexity in other species

2.2.2.1 Mammals

A number of species also produce vocalisations that show some evidence of hierarchical and combinatorial structure, and which may be therefore considered complex in something like the same way as speech and I discuss some examples here. In mammals, several species of **gibbon** sing loud and long songs which reuse a finite number of acoustically distinct elements known as notes (Geissmann, 2000). The function (or functions) of these songs has not been definitively established, but it seems likely that they serve as some form of mate attraction and possibly territorial advertisement (i.e. they are sexually selected), and the complex duets that males and females engage in may serve to strengthen pair bonds (Geissmann, 2000). The function of the complexity of gibbon song is even less clear, but similar arguments as for the function of complexity in bird song (discussed in detail in section 2.2.2.2) may also apply to gibbons (Clarke et al., 2006).

Humpback whales (*Megaptera novaeangliae*) also sing long and varied songs during their breeding season (Payne and McVay, 1971). These songs are again composed of a series of discrete vocal elements, termed units. These units are combined into what appear to be distinct and recognisable patterns which are reused in songs in apparently complex ways (Payne and McVay, 1971). Suzuki et al. (2006) performed a series of sophisticated information theoretic analyses on a corpus of recorded humpback songs and found evidence for hierarchical patterning and long range dependencies suggestive of a sophisticated syntactic structure. Specifically, they argue that the statistics they calculate from the corpus imply that no Markov model of order less than the length of the song itself can fully capture song structure and that therefore a hierarchical structure of the type proposed by Payne and McVay (1971) is justified[1]. The function of these songs has again yet to be definitively established, but as songs are most commonly heard during the breeding season the most likely proposed function of song complexity is again sexual selection.

A recent study has shown that male **mice** (*Mus musculus*) sing ultrasonic ‘songs’ composed of several syllable types which are organised into phrases which are reused in song (Holy and Guo, 2005). The study demonstrates that the transitions between syllables are not fixed and show some relatively sophisticated temporal sequencing. These songs are produced by males when they encounter females or female pheromones, so the likely function of these songs is once more mate attraction.

2.2.2.2 Birds

The songs of the oscine passerines, or **songbirds**, are probably the most widely recognised non-human complex vocal signals. There is a massive variety of song behaviour in this group with the songs of many species being described as complex according to some metric or other. In this section I provide a brief overview of some of the general

---

[1] Of course, it may be the case that the animals do just memorize several complete songs and do not recombine song elements in the ways that Suzuki et al.’s (2006) analyses suggest. However, until we are able to perform detailed neuroanatomical studies of live whales in the process of singing, or at least acquire a better understanding of the neural structures underpinning whale song, I believe that their conclusions seems reasonable. For further details of their specific techniques the reader is directed to the original paper.

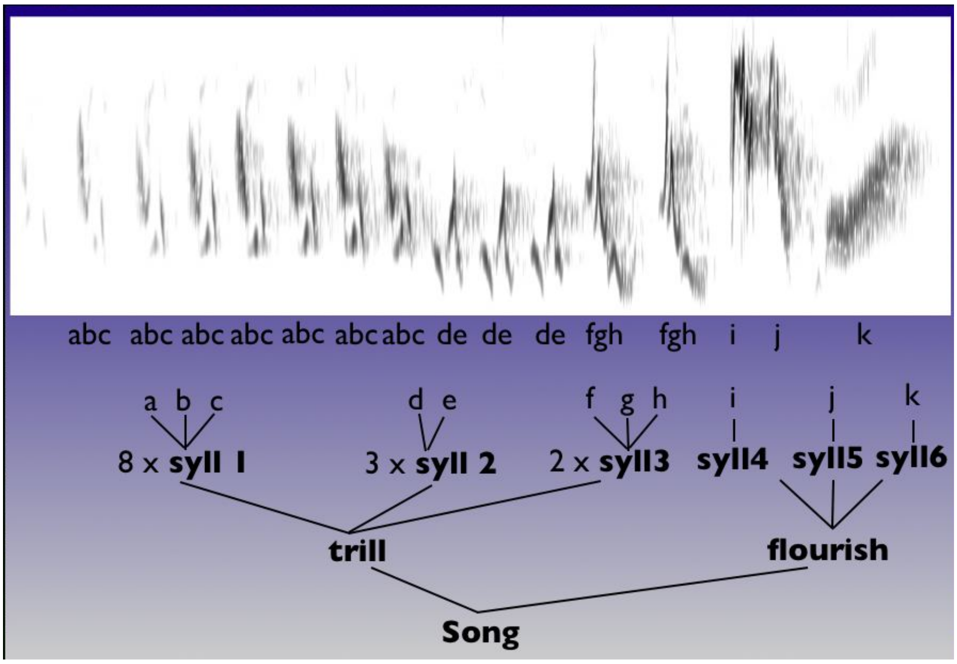


Figure 2.1: A sonogram recording of a single chaffinch song. Note the apparent hierarchy in which individual song elements (marked as characters) are composed into recurring syllables which are then themselves composed into the higher order phrases identified as the trill and flourish. Figure by Robert Lachlan, used with permission.

features of song complexity in several species, and discuss proposed functions of this complexity, for more details see the reviews in (Catchpole and Slater, 1995; Marler and Slabbekoorn, 2004).

In most songbird species song is composed of acoustically distinct song elements. These elements are composed into syllables (or notes) which appear to be the smallest discrete unit of song production; if a bird is interrupted when singing it will finish the current syllable before stopping (Cynx, 1990). Further evidence of the fact that syllables are a meaningful unit of song is that several species, such as the Bengalese finch (*Lonchura striata* var. *domestica*) vary the ordering of song syllables in different songs according to an apparent song ‘syntax’ (Okanoya, 2002), and female canaries (*Serinus canaria*) produce distinct responses to songs composed of different syllables in male song (Vallet and Kreutzer, 1995).

Syllables may therefore be analogous to phonemes in speech in that they are physically continuous but perceptually discrete and are the basic unit of song production

(Doupe and Kuhl, 1999). Syllables commonly occur together in distinct ‘phrases’, and a full ‘song’ is then composed of these phrases (Catchpole and Slater, 1995). Figure 2.1 shows an analysis of a single chaffinch (*Fringilla coelebs*) song showing a purported hierarchical structure. Note that song in this species is composed of two distinct phrases which are on a timescale longer than that of the individual syllable; the trill and flourish. Just as syllables can be rearranged and combined differently, male chaffinches can vary the length of these two phrases independently and sometimes sing songs that do not include the terminal flourish (Leitao et al., 2004). Leitao and Riebel (2003) also show that male and female chaffinches respond to song differently according to the relative lengths of these two phrases; males responded more to songs with longer trills and shorter flourishes, while females preferred songs with longer flourishes. This suggests that the sexes perceive these components of song differently. Phrases may therefore be analogous to words in human speech, in that they are composed of the basic units of vocal production but different ordering or length of the phrase may change the communicative function of the signal.

Within this general pattern of song there is a great deal of further variation found in different species that researchers frequently describe as complex. These variations can be grouped into two main types: song and syllable repertoires (Collins, 2004).

In many species a bird sings several different song types but the individual song types do not vary much. The set of song types a bird produces is known as its ‘repertoire’. Repertoire sizes vary widely in size from 1-6 in the chaffinch to thousands in the brown thrasher (Catchpole and Slater, 1995).

Another manifestation of song complexity is in the sequencing of song syllables. Although several species, such as the zebra finch and the chaffinch, sing songs with fairly fixed syllable ordering which Okanoya (2004) describes as ‘linear’, other species, such as the nightingale, starling and the Bengalese finch sing songs with markedly less deterministic syllable ordering. Okanoya (2004) argues that syllable ordering of Bengalese finch song can be characterised as having finite state syntax.

But what is the function of this apparent complexity? In most species it is only the males who sing, though in some species females also sing and in some males and females sing together in duets (Catchpole and Slater, 1995). Male song appears to serve two main functions; mate attraction and territory defense (Catchpole and Slater, 1995; Collins, 2004). In many species females show a clear preference for the ‘complex’ features of song described above. For example, females in many species appear to prefer males with larger song repertoires. Okanoya (2004) shows that female Bengalese finches display a preference for song with more complex syntactic structure and so argues that mate attraction is again the function of complexity in this species. (I will, however, consider an alternative hypothesis for the origin of song complexity in this species in detail in chapter 4.) A large song or syllable repertoire may also help in territory defense as various studies have shown that males are less likely to intrude into territories where males have larger repertoires (Collins, 2004).

2.2.3 Summary

> primeval man, or rather some early progenitor of man, probably first used his voice in producing true musical cadences, that is in singing, as do some of the gibbon-apes at the present day; and we may conclude from a widely-spread analogy, that this power would 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, 1879)

Other species also produce vocal and non-vocal signals (e.g. many species of insect) that may be described as complex in some way, but the species I have discussed here represent the clearest evidence for vocal complexity in non-human animals. We have seen that several species show evidence of hierarchical patterning and combinatorial structure that does seem analogous with speech. However, it appears that no other species uses this vocal productivity to convey semantic information in a similar manner to humans. Instead, in almost every case, vocal complexity is thought to be associated with sexual selection. It therefore seems like a reasonable hypothesis, as Darwin suggests in the quote above, that complex vocalisation may have evolved in humans in a sexual context before being coopted for the information sharing purpose it serves today.

2.3 Cultural transmission

2.3.1 The cultural transmission of human vocal signals

While there has been a long and ongoing debate about just what innate resources humans bring to language acquisition, there is no doubt that many features of a particular language, such as its lexicon, are acquired by cultural transmission. There is also no doubt that there is a very significant genetic contribution (e.g. Stromswold, 2001). In order for a child to be able to acquire his native language he must be capable of imitating the sounds of other people. The ability to acquire vocalisations by imitation is known by biologists as *vocal learning*. Janik and Slater (1997) define vocal learning as occurring when "vocalizations themselves are modified in form as a result of experience with those of other individuals". They contrast this with *contextual learning* (also referred to as *auditory learning* by, e.g., Jarvis, 2004b) which occurs when learning affects the usage and comprehension of sounds, such as a dog learning to sit when hearing the word "sit". Clearly humans rely on vocal learning to develop a normal language faculty, and are able to acquire new vocalisations by imitation throughout life.

2.3.2 The consequences of cultural transmission in human language

Vocal learning in humans is a pre-requisite for the cultural transmission of language as we know it[2]. As noted in the previous chapter, several researchers have argued that the fact that some features of language are transmitted culturally rather than genetically is not a simple difference in the mechanism of transmission, but that cultural transmission adds another complex dynamic on top of the existing dynamics of genetic evolution and individual development (e.g. Deacon, 1997; Kirby, 1999).

---

[2] It is possible that language could be transmitted culturally by a medium other than speech, e.g., sign language, and such a system would not require a vocal learning ability.

Recent theoretical and experimental work has shown that cultural transmission may have some far-reaching consequences for human language, and may make a key contribution to some particularly interesting features of language. Much of this work stems from the observation that the adult speakers who produce the linguistic data a child learns its language from, acquired language by the same mechanism themselves. This process has been termed *iterated learning*, and several studies have shown that this process may play a key role in the evolution of some features of language (e.g. Kirby, 2001; Brighton, 2002; Smith et al., 2003)

2.3.2.1 The cultural evolution of syntax

One feature of language that has received considerable attention within the framework of iterated learning is the origin of compositional syntax. Compositionality as discussed in these studies is essentially the property that the meaning of a linguistic expression is a function of the meaning of its parts and the way in which they are combined. This is a ubiquitous and apparently adaptive feature in human languages and so might be expected to be a feature derived from natural selection for communicative ability (Pinker and Bloom, 1990). However, a number of formal models have conclusively shown that a compositional communication system can evolve from a non-compositional system in an population of genetically homogenous agents, exclusively as a result of the dynamics of cultural transmission.

Compositional systems only evolve in these models (known as iterated learning models, or ILMs) under certain conditions, and to understand how cultural transmission can have this surprising effect I will briefly discuss the general features of the models that have been shown to effect the emergence of compositionality. An ILM generally has four key components:

- a meaning space
- a signal space
- one or more language learning agents

- one or more language using agents

The task of an ‘adult’, language using agent in the model is assumed to be to teach the learning agent his language by providing the learner with a set of meaning-signal pairs formed according to his grammar[3]. The meanings are selected at random from a meaning space that has some component structure, such coloured objects, e.g. a red triangle or a blue square[4]. The grammar of the first adult agent is initialised such that he uses a random signal to communicate each meaning. The signals are drawn from a signal space which is frequently just a finite set of characters or numbers. The learning agent then uses some learning algorithm to induce his adult language from the set of meaning-signal pairs he was exposed to. Various different learning algorithms are used, but the key characteristic is that the algorithms try to generalise from the training set to produce a parsimonious grammar to use as an adult. The learning agent then replaces adult agent and a new learner is introduced. This cycle is repeated until some stable end state is reached.

In various instantiations of this model, even if the initial signals produced for each of the meanings is entirely random, over time the system evolves to become compositionally structured with parts of the signal being systematically related to parts of the meaning (e.g. Kirby, 2000; Brighton, 2002). The central result of these models is that if a learner is not exposed to all possible meaning-signal pairs, which seems like a reasonable assumption for language, then only compositional systems can survive as they are generalisable from a subset of the input data. The communication system itself adapts to be learnable through this bottleneck.

Kirby (2001) presents another implementation of an ILM which provides a cultural evolutionary explanation of regular and irregular word forms, a phenomenon found in many human languages. The model demonstrates that irregular, i.e. non-compositional, forms can only survive if they are used (i.e. provided to the learning agent) disproportionately often, as they are then less likely to be lost in the bottleneck. In many language irregular forms are indeed frequently found for the most commonly used words, such as the past tenses of the most common English verbs, e.g. ‘be’ and ‘go’.

---

[3] That the learner is explicitly provided with the meaning associated with each signal is a simplifying assumption

[4] The actual implementation of the meaning space is typically just a vector of discrete numbers

Recent work has begun to explore iterated learning in human subjects in the laboratory (Kirby et al., 2008), and initial results indicate that as the theoretical models predict, the ‘languages’ in this study do adapt to be increasingly structured and learnable and that under some conditions compositional structures emerges. Iterated learning may therefore explain some apparently adaptive features of language without resorting to natural selection as the source of this adaptation.

2.3.2.2 Language diversity

The cultural transmission of language also has a clear role to play in language change and diversity. Language learning is not perfect and individual differences in speech production provide a constant source of variation in linguistic behaviour, analogous to the effect of mutation in genetic evolution. This variation can be amplified by geographic separation to give rise to new speech dialects, and by extension, the formation of new languages. There is some debate as to how much of language change is adaptive in some sense, either for the individual speaker or for language itself, or whether aspects of this change are more akin to random drift in genetics (Nettle, 1999; Livingstone, 2002). Wherever one stands on this issue, it is clear that the cultural transmission of language is one of the main sources of linguistic variation. If, as Nettle (1999) argues, language change does serve some adaptive functions then it is possible this would impose a selection pressure for vocal learning as a mechanism that supports this source of variation. I will discuss some analogous proposed functions for vocal learning in supporting song diversity in birds in section 2.3.4 below.

2.3.3 Vocal learning in other species

Vocal learning, by facilitating cultural transmission, may thus contribute to some notable features of language and it turns out that vocal learning is a comparatively rare ability in phylogenetic terms. In the mammals, it has only been clearly demonstrated in humans, cetaceans, seals and bats (Janik and Slater, 1997) though Poole et al. (2005) argue that it may also be found to some extent in elephants. Vocal learning is more

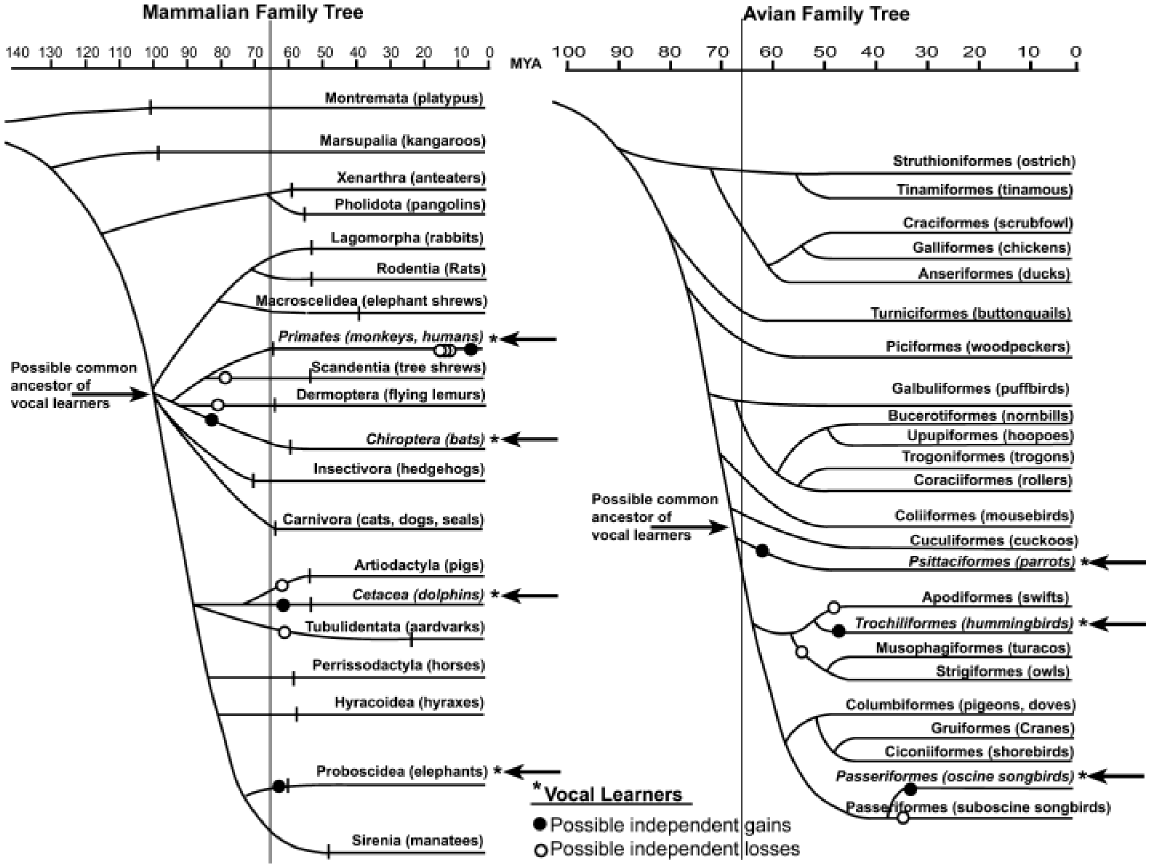


Figure 2.2: Family trees of extant mammalian and avian orders showing the incidence of vocal learning in each of these groups. Figure reprinted from (Jarvis, 2004b) with permission.

widespread in birds, but is still only found in three distinct groups; songbirds, parrots and hummingbirds (Jarvis, 2004a). Notably none of the non-human primates, nor gibbons, despite their song prowess, seem to exhibit this behaviour (Janik and Slater, 1997). Figure 2.2 shows the distribution of vocal learning in the mammals and birds (not including seals).

The disparate distribution of vocal learning suggests that it has independently evolved several times and hence vocal learning in humans is more likely to be analogous with vocal learning in songbirds rather than homologous, i.e. it seems likely that it has evolved independently in our species rather than being inherited from a shared ancestor with all other vocal learning species (Jarvis, 2004a). Accepting that vocal learning has evolved independently in at least some lineages, e.g. in the lineages leading to modern humans and songbirds, allows us to compare hypotheses about the evolution of the behaviour in one species with the other. As we will see in section 2.4 it seems that vocal learning in humans and birds has some striking similarities in developmental terms (Doupe and Kuhl, 1999), and possibly in its neural implementation (Jarvis, 2004b). This suggests that there may be strong epigenetic constraints on the evolution of this phenomenon which makes comparison between the trait in these two lineages even more powerful. In this section I briefly review evidence for vocal learning in other species to see if there are any general trends.

2.3.3.1 Mammals

Several species of **bats** seem to use vocal learning in a number of ways. Greater spear-nosed bats (*Phyllostomus hastatus*) use vocally learned calls to identify themselves as members of a particular foraging group (Boughman, 1998). Greater horseshoe bats (*Rhinolophus ferrumequinum*) learn the frequency of their echolocation call from their mother (Janik and Slater, 1997), and mother lesser spear-nosed bats (*Phyllostomus discolor*) exchange calls with their infants to help locate each other (Janik and Slater, 1997). A common function of vocal learning in this group then appears to be for individual recognition, frequently between mother and infant.

A number of **cetacean** species are also capable of learning novel vocalisations, two of the the best studied species are the humpback whale (*Megaptera novaeangliae*) and the bottlenose dolphin (*Tursiops truncatus*). As discussed earlier, male humpbacks sing long and complex songs in the breeding season. These songs change throughout the singing season, but generally all males in a group sing the same song providing clear evidence of imitation (Janik and Slater, 1997). The function of song imitation and variation in whales is not yet well understood, but it is possible that synchronised singing and song variation serves to stimulate and attract females (Janik and Slater, 1997). Dolphins are capable of imitating pulsed sounds and whistles (Tyack and Sayigh, 1997). Individuals produce distinctive whistles that may serve to identify them to conspecifics, termed ‘signature’ whistles, so vocal learning may provide a means of individual and group recognition.

One of the most convincing demonstrations of vocal learning in **seals** was the harbour seal (*Phoca vitulina*) known as 'Hoover'. Hoover spontaneously imitated a number of phrases clearly recognisable as English, including: "hey there" and "come over here" (Ralls et al., 1985). Evidence of vocal learning in the wild in this group has generally only come from studies demonstrating the differences in vocalisations in geographically separate populations (Janik and Slater, 1997). Song in seals is often associated with mate attraction so it seems likely that vocal learning in this group may once again be associated with sexual selection (Janik and Slater, 1997).

One recent study (Poole et al., 2005) claims that **elephants** may also be capable of vocal learning. They provide some statistical analyses suggesting that two African elephants (*Loxodonta africana*) may have acquired calls by vocal imitation. The first, a female known as Mlaika, appears to mimic the sound of trucks driving nearby. The second case was in a male known as Calimero who spent 18 months with two Asian elephants (*Elephas maximus*) and who produces a chirp call typical of Asian elephants, but not typically made by African elephants. Given that this behaviour is found in both sexes, the authors argue that social bonding or individual or group recognition are the most likely candidate functions. More empirical work is required though to definitively establish this phenomenon in this species and to investigate possible functions.

2.3.3.2 Birds

The imitative ability of **parrots** is one of the most obvious demonstrations of vocal learning ability outside of humans. Some **hummingbirds** have also been shown to acquire their songs by vocal learning (Jarvis et al., 2000). Undoubtedly though, the best studied group of vocal learning birds are the **songbirds**. Every songbird species studied to date learns its song (Catchpole and Slater, 1995). The ontogeny of song learning varies between species, but there are general patterns and I discuss some of these in section 2.4 below. There have been a large number of functions proposed for song learning in the songbirds and I discuss several of these in the following section.

2.3.4 Proposed functions for vocal learning in bird song

The fact that every oscine bird studied to date learns its song makes it likely that vocal learning evolved early in the history of the group, possibly in a shared ancestor with parrots up to 100 million years ago (Kroodsma, 2004). We are therefore unlikely to ever be able to determine the ecological conditions that led to the original evolution of this behaviour. However, it seems plausible that as song learning is so widespread there must at least have been selection favouring its maintenance (Lachlan and Slater, 1999), especially given that there are a number of possible evolutionary costs imposed by learning, for example:

- Song learning requires a substantial investment in time to develop normal adult song.

- As outlined below, song learning depends on a complex set of neural systems and these may be costly to develop and maintain, both over the lifetime of an individual and in phylogenetic terms.

- Learning introduces the risk that suitable models are not available; a bird may acquire heterospecific song or simply fail to find a tutor at all.

- To the extent that learning facilitates more varied song, this may lead to increased predation as predators are less likely to habituate to such variation than to a stereotyped universal song (Jarvis, 2004b).

The fact that these various costs of one form or another must be paid in order to develop normal song suggests that there must at least have been selection for the maintenance of song learning behaviour in the more recent evolutionary past. Jarvis (2004a) argues that one plausible hypothesis is that vocal learning emerged in the ancestor of all current avian vocal learners and was then independently lost in four separate lineages, while being retained in the parrots, hummingbirds and oscines.

Several hypotheses have been proposed that attempt to explain either current functions of song learning, and hence its maintenance, or possible reasons for its original evolution. As this question is one of the central issues of this thesis, in this section I provide an overview of the most important hypotheses, and discuss issues with them.

Advantages over genetic transmission Nowicki et al. (1998) note that one traditional view of the advantages of song learning (e.g. Nottebohm, 1972) is that it may be difficult to encode the motor information required for complex song genetically and so accurate transmission requires vocal imitation. However similarly complex behaviours in many animals, such as gibbon song, do appear to be transmitted largely genetically and some song learners such as zebra finches have comparatively simple stereotyped songs.

Individual and group recognition As has been argued for bats and dolphins, vocal learning may facilitate the acquisition of distinctive songs that allow individual recognition. Song sharing by birds on neighbouring territories has been observed in several species and may serve to maintain territory boundaries (Collins, 2004). However, as Nottebohm (1972) argues, accurate learning will generally lead to individuals sounding more alike. He also notes that normal inter-individual differences in physiology etc. may often serve well enough at individual recognition.

Habitat matching Birds of one particular species may live in a wide range of habitats and it may be that some song types are more suited to one environment than another, perhaps due to local differences in the transmission of sound. Hansen (1979) first argued that song learning may have evolved as a way of matching song to habitat. There is evidence for habitat matching in some some species, for example Tubaro and Segura (1994) found a relationship between habitat and several features of the trill of Rufous-collared sparrow (*Zonotrichia capensis*) song. However, in these studies the effect is not always that strong and can be disputed in boundary areas and there are a number of complicating factors, such as physiological differences that are also related to habitat (Catchpole and Slater, 1995). This effect has also been looked for and not found in other species, e.g. chaffinches (Williams and Slater, 1993).

Genetic adaptation As Nottebohm (1972) argues, panmixia (random mating between all members of a species) between populations may slow down the development of local adaptations and break up locally adaptive gene complexes. It may therefore be advantageous in some circumstances for assortative mating with individuals from the same local population. Local song dialects may help to achieve this, and as dialects are a consequence of, or at least rely on, song learning, this behaviour may be selected as a result. This hypothesis relies on the key assumption that both males and females learn the song of their birth area and later settle there. This assumption has been investigated most thoroughly for the white-crowned sparrow, as discussed in (Catchpole and Slater, 1995), and the evidence is not convincing; some males seem to learn from tutors after dispersal, females injected with male hormones did not sing songs typical of their birth area and females do not appear to prefer males singing the same dialect as their father (references in Catchpole and Slater, 1995, pps. 206-7). This assumption also does not seem to hold for a number of other species.

Sexual selection As song is mainly used in a sexual context, the effect that particular types or variations of song have on potential mates and on potential invaders has been extensively studied. As we saw in section 2.2.2.2, song complexity is generally associated with sexual selection. If song learning is able to increase song complexity in some way then it may be selected for similar reasons as for complex song; primarily in mate attraction. This argument seems most applicable to the acquisition of a large and varied repertoire. As discussed above, females in many species appear to prefer males with larger song repertoires and vocal learning is one mechanism by which a bird could acquire a variety of song types.

Sexual selection was the assumed driver for the evolution of vocal learning by Aoki (1989) in one of the first attempts to formally model the evolution of vocal learning in bird song (and in any other species as far as I am aware). Aoki assumes that in a population with an innately determined song (the presumed ancestral state of the oscines), a genetic mutation occurs which results in a mutant male who is able to invent a variant song that females prefer to the innate song. This variant song can then be transmitted culturally to other mutant males. The female preference is assumed to be genetically determined, and he presents two models, one where the gene encoding female preference is already present and fixed in the population and a second two locus model where the production and preference genes coevolve. Aoki's population genetic model shows that with these assumptions the learning allele can invade the population in some specific circumstances.

Aoki's model is interesting and provides insights into the dynamics of the evolution of a culturally transmitted trait, but his assumption that the mutant learning male's song is "serendipitously" preferable to females does not suggest any functional role for vocal learning. Aoki's argument for the evolution of vocal learning is then that, essentially by chance, song acquired culturally is preferable to song transmitted genetically, this relies on the assumption that the preferred song cannot be encoded genetically. His model also relies on the biologically unrealistic assumption that males acquire their songs from their fathers and not unrelated males, as is the case in many species (Lachlan and Servedio, 2004).

A recent proposal for the function of song learning in a sexual selection context is that it allows females to judge a male's quality by how well he has acquired conspecific song (e.g. Nowicki et al., 1998). I look in detail at this hypothesis in chapter 5.

The 'cultural trap' hypothesis Lachlan and Slater (1999) argue that while several of these hypotheses seem plausible and empirically supported for some species, none can convincingly apply to all. For example the zebra finch's song is fairly simple and stereotyped across individuals and groups (at least so far as we can tell), and so it is not clear why, given the apparent costs of learning, it should still be transmitted culturally. Lachlan and Slater (1999) suggest an interesting hypothesis for the evolutionary maintenance of vocal learning that seems to apply fairly widely to vocal learning songbird species, and may also apply to vocal learning in other species. Their 'cultural trap hypothesis' suggests that once a song learning system was in place, the dynamics of the interaction of genes and culture may be sufficient to explain the maintenance song learning without any effect on fitness. They demonstrate this effect with a computational model which Lachlan and Feldman (2003) then later generalised and investigated further with a population genetic model.

In the first of these papers they introduce a useful theoretical framework for investi-

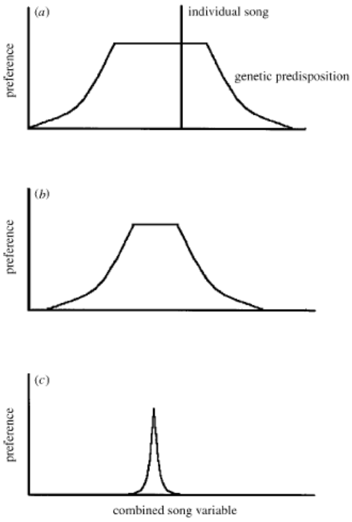


Figure 2.3: A representation of the genetic song filter as used in (Lachlan and Slater, 1999). The y-axis represents the preference for a particular value on the single song dimension, represented by the x-axis. (a) shows an example song being picked. (a)-(c) demonstrate learning disappearing as the filter narrows. This figure is taken from (Lachlan and Slater, 1999, p. 702).

gating the question of the maintenance of song learning, which I will briefly present here.

Song variation is modelled as a one-dimensional measure, which they term the 'combined song variable'. Song learning involves picking one value on this dimension. This is not done freely as there is some genetic bias to particular values, or ranges of values. This genetic bias can be thought of as a filter on all possible songs which normally ensures that only species-specific song is acquired. The interaction of the genetic preference and the single song dimension is shown in figure 2.3.

Lachlan and Slater (1999) argue that if learning was selected against the filter would progressively narrow until there was no longer scope for any variation and learning would essentially have disappeared, as shown in the diagram. It seems plausible that a single mutation could narrow or widen this filter and so in order to investigate the conditions under which learning is maintained is this model they compare the success of two alleles, a 'wide' allele which causes the filter to be wider than a 'narrow' allele. They assume that both males and females inherit these filters and a female uses her filter to assess her preference for a male's song, while a male uses his to acquire his adult song.

Defining the model this way implies that males with wide alleles may acquire a song outside the filters of females with narrow alleles, while females with narrow alleles may not recognise some males. Both of these eventualities will be selected against and so they argue that there should be antagonistic selection against both alleles if they coexist in a population. However they argue that a population in which the wide allele is at fixation is more stable to invasion by the narrow allele than the reverse situation. When the wide allele is common, any individual born with the narrow allele is likely to be unable to learn some of population's songs and will be selected against. When the narrow allele is common, on the other hand, an individual born with the wide allele *will* be able to learn songs from singers with narrow alleles. This imbalance means that a population fixated on the wide allele is likely to be more stable to invasion that a population fixated on the narrow allele. They demonstrate that this argument holds under a wide range of conditions, both with a computational agent-based model in (Lachlan and Slater, 1999) and with a population genetic gene-culture co-evolutionary model in (Lachlan and Feldman, 2003).

This hypothesis is interesting as it does not rely on any particular ecological conditions, it is a consequence of the interaction between genetic and cultural evolution, and can therefore apply wherever their assumptions about the genetically-based song filter and the mechanism of cultural transmission and mutation seem reasonable. Although the hypothesis is mainly concerned with the maintenance of song learning, the authors note that a more general interpretation of the model could construe the wide allele as a learning allele and the narrow allele as a non-learning allele, and the same arguments would apply. The model may then, with some caveats, be seen as a possible explanation for the *origin* of song learning as well as its maintenance.

Surveying the hypotheses As we have seen, there are a number of plausible advantages to song learning, but no single functional hypothesis seems to apply to all species. It may be that there is no single explanation and that song learning may have evolved or been maintained in each species due to different functions but this is a rather unparsimonious conclusion, and we will look at one possible unifying hypothesis in detail in chapter 5.

2.3.5 Summary

Current advantages of vocal learning ability for humans seem clear; it is hard to imagine how language could function if we were not able to learn to produce new words. However, selection pressures favouring its maintenance may not necessarily explain its origins. It is possible that we acquired our vocal learning abilities before this pressure for expressibility existed. In this brief review of vocal learning in other animals we have seen that there are several other possible functions of vocal learning, and it is possible that this ability originally evolved in humans to serve similar functions.

2.3.6 The co-occurrence of cultural transmission and complex vocalisation

Having reviewed the distribution of these two key features of speech, there is only clear evidence of these two behaviours co-existing in two other groups; cetaceans and the songbirds. An understanding of the evolution of the communication systems of each of these groups is therefore likely to be very relevant to study of the evolution of language. In this thesis I will concentrate on bird song as a model system, largely due to the wealth of empirical data available from a wide range of perspectives.

2.4 Bird song as a model of speech

> The sounds uttered by birds offer in several respects the nearest analogy to language, for all the members of the same species utter the same instinctive cries expressive of their emotions; and all the kinds that sing, exert their power instinctively; but the actual song, and even the call notes, are learnt from their parents or foster-parents. (Darwin, 1879)

We have seen that bird song clearly displays the two characteristics of human speech that concern this thesis. The rest of this thesis uses bird song as the main model for the evolution of these features. As well as sharing these two features of language there are a number of other striking parallels between song and speech some of which I discuss here.

2.4.1 Physiological and neural parallels

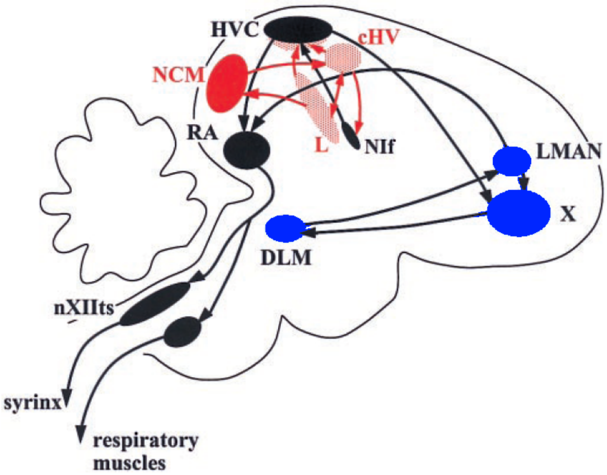


Figure 2.4: Schematic diagram of the main song control centres in the songbird brain. Nuclei from the motor pathway are shown in black, and nuclei from the learning pathway are in blue. Figure after (Marler and Doupe, 2000)

Both speech and song are complex acoustic signals generated by air flowing over specialised organs which are controlled by sophisticated neural systems. Although the mammalian and avian vocal systems and the neural systems underlying them are clearly evolutionarily distant there are some interesting potential parallels in the physiology underlying song and speech production.

Songbirds sing using a specialised vocal organ, the syrinx, which is analogous to the human larynx. The syrinx contains membranes which vibrate and generate sound waves when air from the lungs is forced over them. These sound waves are then altered as the acoustic signal passes through the vocal tract and out through the beak. The syrinx is controlled by a specialised neural pathway depicted schematically in figure 2.4, the regions highlighted are the main song control nuclei (anatomically discrete clusters of neurons) in a songbird's brain, and the arrows between them represent neural axons projecting to other nuclei, forming distinct neural pathways. The nuclei in blue represent the main motor pathway which runs from the higher vocal centre (HVC) to the robust nucleus of the archistratum (RA). From the RA there is a projection to the tracheosyringeal portion of the hypoglossal nucleus (nXIIts), which then in turn controls sound production in the syrinx itself.

The HVC and RA nuclei control song production, but there is a distinct pathway for song learning as well, identified in black in the diagram. Area X receives a projection from the HVC, which is indirectly (via the dorsolateral nucleus in the thalamus (DLM)) connected to the LMAN region which then projects to the RA song control nucleus. There are thus 2 distinct pathways from the main HVC nucleus to the RA before song is produced, one is direct (HVC - RA), and is known as the posterior pathway, and the other is indirect (HVC - X - DLM - LMAN - RA) and is known as the anterior pathway. This is known as the ‘recursive loop’ (Marler and Doupe, 2000) and may allow a bird to compare or recognise song structures stored in different areas of the brain, which may explain some features of the song learning process discussed below.

In human speech, the signal is again generated as air from expiration passes over vocal folds in the larynx. Components of this waveform are then altered as they pass through the rest of the vocal tract and some components are ‘filtered’ out leaving peaks of acoustic energy known as formants (Fitch, 2000). Formants constitute much of the vocal information carrying capacity of speech, for example vowels are characterised by fairly constant formant frequencies while consonants are composed of rapidly changing format frequencies (Doupe and Kuhl, 1999). Doupe and Kuhl (1999) argue that the avian vocal tract may play a similar role in forming the characteristic sounds of song.

We know rather less about the details of the neural systems controlling speech production, but a number of authors argue that there may be some functional similarities. For example, Okanoya and Merker (2007) speculate that Broca’s area, a region known to play an important role in language abilities, may be comparable with the HVC as it receives input from the auditory cortex and then has pathways onwards to the face motor cortex which controls vocalisations and is thus analogous to the posterior pathway in birds. They also suggest that there may be similar secondary anterior pathway from Broca’s area to the motor cortex via the basal ganglia and anterior cingulate.

Jarvis (2004b) also argues that there may be functional similarities in the neural implementation of vocal learning in humans and birds. He presents a detailed comparative study showing that the three groups of vocal learning birds (songbirds, hummingbirds and parrots) have seven similar but not identical cerebral vocal nuclei and functionally similar neural pathways that facilitate vocal learning. He argues that vocal learning has evolved independently in each of these groups and so the similarities in neural implementation suggest that there are strong constraints on the evolution of vocal learning in the basic neural network of the vertebrate brain. He agrees with Okanoya in postulating that humans may also have anterior and posterior pathways which underly our vocal learning abilities.

2.4.2 Developmental parallels

As the quote from Darwin above suggests, the fact that both language and song are acquired from conspecifics is one of the most frequently noted similarities between the two behaviours. Interestingly, the ontogeny of song and speech also share some characteristics some of which I review briefly here. The reader is referred to (Doupe and Kuhl, 1999) for an extended comparison of common themes in the development of speech and song.

The interaction of auditory experience and innate biases Both humans and birds are born with genetic biases to conspecific sounds but require exposure to conspecific vocalisations to develop normal communication abilities. As an example, a song sparrow that is raised in isolation and not exposed to any conspecific songs will still produce song as an adult. This ‘isolate’ song is clearly abnormal but retains some recognisable (for a human anyway) features of song sparrow song (Marler, 2004). This is a clear indication that there is a genetic contribution to song structure, however to develop normal song a bird must also hear conspecifics. There are a few reported incidences of human children being raised with little or no exposure to human speech, with significant deleterious effects on language development (Doupe and Kuhl, 1999). While humans need exposure to speech to develop normally, infants also appear to discriminate human speech sounds from other sounds from a very early age which suggests that there is also some innate component to our vocalisations (Doupe and

Kuhl, 1999). To what extent these biases are specific to humans rather than shared with other primates is the subject of ongoing research.

Immature vocalisations The vocalisations of both babies and nestlings are easily differentiable from adult sounds. In birds early vocalisations are noticeably quieter and contain sounds that adult birds do not produce, this stage is known as subsong. As the infant matures subsong turns into plastic song which contains elements learned from conspecifics, but with less fixed ordering and patterning than seen in adults. Generally by the end of a bird's first spring, song will have crystallised into adult song and will not change much more throughout a bird's life (Catchpole and Slater, 1995). There are of course exceptions to this general model, and some species can modify their vocalisations later in life. Human infants also go through various distinct phases of immature vocalisation, from initial use of sounds that do not appear in human speech, to babbling by the age of 7 months and the first recognisable words by around 10-12 months (Doupe and Kuhl, 1999).

The role of auditory feedback Both humans and birds require auditory feedback to develop normal vocalisations; neither deaf babies nor deaf birds develop normal vocal abilities. For example male song and swamp sparrows who were exposed to normal conspecific song but were then deafened before song production started produced amorphous and very abnormal song (Marler and Doupe, 2000). Clearly the inability to hear early vocalisations severely affected the development of song production. Studies of children who become deaf before puberty show that humans also need to hear themselves in order to develop normal language abilities (Doupe and Kuhl, 1999). This dependence on auditory feedback lessens for both groups in adulthood. Adult humans who become deaf later in life do show deterioration of speech but can often still be understood. Some birds deafened in adulthood show very little degradation in song (Doupe and Kuhl, 1999), while other species require auditory feedback throughout adulthood to maintain normal song (Okanoya and Yamaguchi, 1997).

Critical periods The capacity to acquire accurate copies of model sounds does not last throughout the lifetime of birds or humans. In both it appears that if exposure to conspecific sounds does not occur before a fairly specific time (puberty in children and sexual maturity in birds), later exposure is not sufficient to produce normal vocalisation (Doupe and Kuhl, 1999). Clearly humans are able to acquire new words throughout their lives, but the existence of 'foreign accents' is clear evidence for a reduced capacity for accurate vocal imitation.

2.4.3 Genetic similarities?

The physiological and developmental mechanisms discussed above must be underpinned by genetics but we still know comparatively little about the molecular basis of vocalisation in general and vocal learning in particular. Nevertheless, some recent work studying one particular gene implicated in human speech suggests that there may be some parallels even at the molecular level in vocal learning species. Lai et al. (2001) isolated a single gene on human chromosome 7, known as FOXP2, which is implicated in speech production. Individuals with a mutation in this gene have significant impairments to language comprehension and speech production (Marcus and Fisher, 2003). Work comparing the DNA sequence of this gene in humans and in chimpanzees has established that the human version appears to be functionally identical in all human populations, but has two amino acid changes compared to chimpanzees (Enard et al., 2002). There is also evidence that this gene has recently been subject to natural selection (Enard et al., 2002). FOXP2 is thus the first gene to have been identified as playing a key role in human speech and appears to have been subject to recent selection.

Songbirds have a homolog of this gene which is known as FoxP2, and the protein sequence of this gene is 98% similar to the human version (Haesler et al., 2004). This gene has also been demonstrated to play a role in song development, Haesler et al. (2004) show that expression of this gene in Area X, an important song control nucleus, in zebra finches is highest during 30-45 days after hatching when vocal learning occurs in this species, they therefore conclude that FoxP2 in songbirds is associated with vocal plasticity. In more recent work, Haesler et al. (2007) provide more evidence for this hypothesis and demonstrate that experimentally reducing levels of FoxP2 expression in infant zebra finches resulted in the birds acquiring incomplete and inaccurate models of their tutor songs and these effects lasted into adulthood. Teramitsu et al. (2004) also argue that the expression pattern of FOXP2 in humans and FoxP2 in birds are strikingly similar.

These results are interesting and suggest that there may be some low level molecular similarities between vocal learning abilities in songbirds and humans. However vocalisation and vocal learning abilities are almost certainly underpinned by a large number of genes and further work is required to demonstrate whether this parallel is indicative of a general genetic similarities, or if this early result is misleading.

2.4.4 Summary

While several of these parallels have yet to be fully empirically substantiated, in sum they do seem to provide clear support for the hypothesis suggested by Jarvis (2004b) that there are strong constraints on the evolution of complex vocal learning abilities in vertebrates, that result in similar evolutionary solutions to the problem of vocal imitation at a number of levels. This suggests these traits may also have similar evolutionary histories and so hypotheses proposed for the evolution of complex vocal learning in birds may be applied to the evolution of these abilities in humans.

These parallels are very interesting but there are also a number of important dissimilarities between song and speech. Perhaps the most obvious is that in songbirds generally males do most of the singing. While there are exceptions to this rule (Riebel, 2003), in humans there is no such dimorphism and there is little difference in vocal behaviour between males and females, besides differences in voice pitch. As discussed earlier, songbirds also do not use song to communicate arbitrary information to each other, although complexity in song structure may provide information about the singer to a potential mate.

2.5 Conclusions

This chapter has argued that two components of the human language faculty; complex vocalisation and cultural transmission, make a significant contribution to the nature of human language. In particular the evolution of complex vocal learning abilities is a necessary precursor of a system of iterated learning for language, which I have argued may explain some important observed features of language. Both of these features are shared to some extent with other species and so hypotheses that attempt to explain the evolution of these traits in the various species may also apply to the evolution of these traits in humans. We have seen that other species use these abilities in a number of ways that seem quite different from their current functions in humans and so it is possible that these features evolved in humans for functions other than that they are used for now. In particular we have seen that vocal complexity and vocal learning are frequently associated with mate attraction and other sexually selected functions, and this supports old (e.g. Darwin, 1879) and more recent (e.g. Mithen, 2006) suggestions that early human communication may have been similar to song, perhaps as we see in gibbons today.

The rest of this thesis investigates several issues in the evolution of complex vocal learning abilities using bird song. As I have argued in this chapter, bird song is a very useful model of these abilities in humans because of a number of striking parallels in the behaviours. There are also considerably more empirical results pertaining to the physiology, development and proposed evolutionary function of song.

Studying the evolution of these abilities is a significant theoretical challenge as a result of the non-trivial and sometimes counter-intuitive interactions between genetic evolution, individual development and the resulting cultural evolutionary processes. Previous work in this vein has used formal models to make the contribution of each of these components explicit and this is the approach taken in the rest of this thesis.

Chapter 3

A framework for modelling the evolution of vocal learning in bird song

In this chapter I describe a simple computational model of song learning inspired by the most widely accepted model of song learning, the 'auditory template model' (ATM). The model includes an explicit formalisation of the various components posited by the ATM. I then use this model of learning in a computational agent-based model to investigate some conditions under which song perception and production can come to be increasingly influenced by cultural transmission. In particular I investigate the effects of the reliability of the cultural environment and the stage of life at which communicative behaviour is required. The models demonstrate that song learning is only likely to evolve under certain environmental and life history constraints, regardless of any proposed function for vocal learning[1].

---

[1] A paper including much of the content of this chapter has been published as (Ritchie and Kirby, 2006) and is included in the appendix. This paper is jointly authored with Simon Kirby with whom I discussed the overall design of the model, but otherwise the implementation, experiments and analysis are my own work.

3.1 The auditory template model of song learning

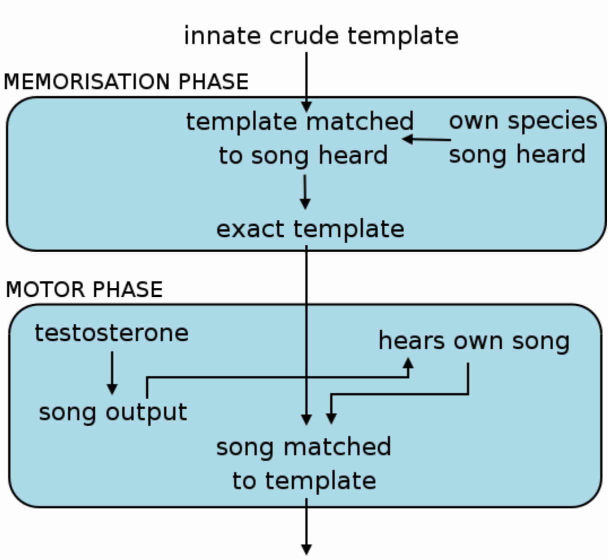


Figure 3.1: The auditory template model of song development. Figure after (Catchpole and Slater, 1995)

As discussed in the previous chapter, while the development of song varies significantly between different songbird species, there does appear to be some general patterns. In attempting to capture the general features, researchers have developed what is known as the 'auditory template model', depicted graphically in figure 3.1 and describe in details in (Catchpole and Slater, 1995). This model assumes that birds are born with an genetically specified 'crude' template which defines the approximate characteristics of their species song. The bird then goes though an initial sensitive period during which it will hear songs from conspecifics, but may also hear the songs of other species, and memorises songs which match its crude template, thereby refining the crude template to an exact template which constitutes their memorised model of song.

A bird does not normally not begin to produce song-like vocalisations (as opposed to calls which are produced from birth) until its first spring when testosterone begins to circulate in its body. The bird then goes through a process of matching its own vocalisations to the template it has memorised. At first the song is very variable and shows little in common with its species song, this stage is referred to as 'subsong', but gradually with practice the bird learns to produce an increasingly accurate version of its template and the song 'crystallises' into the species typical song. After the song has crystallised it will not vary much again, and further exposure to other songs will not alter the birds song.

This model is supported by a lot of empirical evidence some of which is discussed in section 2.4 in the previous chapter. I take this as a plausible model of song learning which informs the design of the models presented in this chapter.

3.2 A simple computational implementation of the auditory template model

The ATM posits two distinct phases to song learning; an early memorisation phase in which songs that are heard as an infant which are recognised as conspecific by an innate ‘crude template’ are memorised, and a later motor phase when song production is trained to produce songs that match the learned template. I take this model as inspiration and develop a computational model of the two proposed stages of learning in bird song, described in the following sections.

3.2.1 Phase 1: Observational learning

As discussed in section 2.2.2.2 many songbird’s songs are composed of distinct elements, quite how elements are defined varies from species to species, but for simplicity here I refer to an atomic song element as a ‘note’ (this may be analogous to a phoneme in speech). I model a song as a string of such notes, represented here by single alphabetic characters in a similar way to (Sasahara and Ikegami, 2003; 2004). I assume that the notes are fixed and identical for every agent, and the number of notes is defined as a parameter n_{notes} (for the models described below this parameter was set to 6).

Under this model then, the only parameter of song that is learned by an infant bird is the sequence of song elements. While this seems a reasonable model for some species such as the Bengalese finch (*Lonchura striata* var. *domestica*) (Okanoya, 2004) and the white-crowned sparrow (*Zonotrichia leucophrys*) (Rose et al., 2004) which appear to have fairly fixed song elements across species, while different individuals and populations vary in the sequencing of these elements, this is obviously an oversimplification. Many species (including, to some extent, the two just mentioned) also learn the form of song elements from conspecifics, as well as potentially other song parameters such as trill rates and syllable lengths. However, only considering one song parameter makes modelling more tractable and the results clearer while still retaining the key characteristics that there is both a genetic and cultural contribution to adult song structure. I do not believe that the results discussed here would be significantly affected by the particular song parameter that I include in the model. Also, while Cynx (1990) shows that the syllable (as the 'notes' used are intended to model) may be the smallest meaningful unit of song for an individual bird, these elements are physical phenomena that can be themselves be considered to be composed of smaller acoustic units. It may be the case that the acquisition of the form of individual song elements can reasonably be considered as the learning of a sequence of these smaller units which are probably a consequence of song system physiology and may therefore reasonably be assumed to be innate (Okanoya, personal communication). This model of vocal learning as acoustic element sequence acquisition may also therefore apply to the learning of song element form as well as sequence.

To model the memorisation phase of song learning I hypothesise a module which I term the Song Recognition Device (SRD). This is intended to model the auditory biases infant birds show towards conspecific song. I model the SRD as a note transition matrix which defines the transition probabilities between every available note[2]. Each agent is assumed to have 'genes' which code for an innate SRD, this is intended to model the 'crude template' posited by the ATM. An example note transition matrix and the genes that code for it is shown in table 3.1

The SRD transition matrix serves one main purpose; to establish the probability that a given song is one of the bird's own species song. This is done by establishing the average probability of each note transition in the song, as shown in equation 3.1 which defines the preference a given matrix m_x has for a particular song s_y. In this equation n is the number of note transitions in s_y and $m_x(t_i)$ is the entry in m_x for the ith transition

---

[2] While I implement the SRD as a note transition matrix here this component could be modelled in many different ways, e.g. as a neural net with the initial weights specified genetically.

of s_y. For example the preference value the matrix in table 3.1 gives for the song *cab*, which has the transitions *S*-*c*, *c*-*a*, *a*-*b* and *b*-*E*, is $\frac{0.5+0.6+0.5+0.5}{4}$ = 0.525, while the preference for the song *bac* is 0.125. Note that we always include the transition to the first note and from the last note.

$$preference(m_x, s_y) = \frac{\sum_{i=1}^{n} m_x(t_i)}{n} \tag{3.1}$$

Such a matrix can be more or less biased to a particular song-type. If all the probabilities in the matrix are equal then the matrix has no preference to any particular song, while if each row has exactly one high probability transition, the matrix is maximally biased to one particular song. We can measure this bias by calculating the Shannon entropy (Shannon, 1948) for each transition distribution, or row r_x, in the matrix, as shown in equation 3.2. We can then calculate the bias strength of a matrix m_x by summing these values, as shown in equation 3.3. Note that this value will be 0 for a maximally biased matrix and 1 for a matrix with equal probabilities for all transitions. For example the strength of the matrix in table 3.1 is 0.81. This provides us with an explicit value for the strength of an agent's innate bias to a particular song type.

$$entropy(r_x) = \frac{-\sum_{i=0}^{n_{notes}} p_i \; log\;(p_i)}{log(n_{notes})} \tag{3.2}$$

$$strength(m_x) = \frac{\sum_{i=0}^{n_{notes}} entropy(r_i)}{n_{notes}} \tag{3.3}$$

To model the acquisition of the 'exact template', an agent's adult SRD is subject to being altered by songs heard in early life. I model this by 'exposing' each agent to *m* songs (with *m* set to 100 for all results described here) from its environment and deciding whether to add each to a training set of model songs probabilistically according to the preference value of its innate SRD (crude template) using equation 3.1. The note transitions in this set of model songs are then reinforced in the agent's SRD to produce the agent's adult SRD, or 'exact template'. This is implemented by computing the transition probabilities in the training set and creating a second 'learned' matrix, an example of which is shown in table 3.2, for the set of model songs $\{cab, cabcab, cab, abcab, cabcab\}$. This matrix is then merged with the innate matrix to produce the birds adult SRD matrix. The degree to which an agent's SRD is modifiable by songs heard in early life is determined by genes which code for the agents SRD plasticity (SRDP). This will be a value between 0 and 1, with 0 meaning the innate SRD is entirely fixed, and 1 meaning that the agent relies only on its model songs to construct its adult SRD. Each entry in the matrix i, j is calculated according to equation 3.4. An example adult SRD computed with with an SRDP value of 0.5 and the previous example genetic and learned matrices is shown in table 3.3.

$$SRD_{adult}[i,j] = SRDP \cdot SRD_{innate}[i,j] + (1 - SRDP) \cdot SRD_{learned}[i,j] \tag{3.4}$$

	a	b	c	E
S	0.20	0.13	0.47	0.20
a	0.13	0.47	0.20	0.20
b	0.13	0.00	0.40	0.47
c	0.64	0.18	0.09	0.09

0.3	0.2	0.7	0.3	0.2	0.7	0.3	0.3	0.2	0.0	0.6	0.7	0.7	0.2	0.1	0.1

Table 3.1: The genetically specified SRD note transition matrix and the chromosome that codes for it. To calculate the probabilities for each row in the matrix I normalise each block of $n_{notes} + 1$ loci. For these examples n_{notes} is set to 3 meaning that the chromosome will have $(3+1)^2 = 16$ loci (6 notes were used for all simulations described here, and hence chromosomes had 49 loci). The S indicates the start of the song, and the E indicates the end of the song. All values in the tables are shown to 2 decimal places.

	a	b	c	E
S	0.20	0.00	0.80	0.00
a	0.00	1.00	0.00	0.00
b	0.00	0.00	0.38	0.62
c	1.00	0.00	0.00	0.00

Table 3.2: The learned SRD matrix representing the note transition probabilities of the model songs selected as the training set: cab, $cabcab$, cab, $abcab$, $cabcab$

	a	b	c	E
S	0.20	0.07	0.64	0.10
a	0.07	0.74	0.10	0.10
b	0.07	0.00	0.39	0.55
c	0.82	0.09	0.05	0.05

Table 3.3: The adult SRD matrix which integrates the genetic and learned matrices. For this example SRDP is set to 0.5, so, for example, the transition from a-c is calculated as $(0.5 \cdot 0.2) + (0.5 \cdot 0.0) = 0.1$.

3.2.2 Phase 2: Reinforcement learning

The SRD models an agent's sensory biases (or lack thereof) to a particular song-type. We also require a model of song production, which I again implement as a note transition matrix which I term the Song Production Device (SPD). However, for the SPD the transition probabilities determine the probabilities of singing one note after another. Just as for the SRD, an agent has innate biases for its SPD encoded in its genes.

To model plasticity in the production mechanism, I allow the adult SPD to be trained by reinforcement learning using the agent's adult SRD as a critic, using a very simple learning algorithm. This is intended to model the process by which a bird uses its memorised exact template to guide its vocal development. As for the SRD, the degree to which the adult SPD is allowed to be influenced by learning, the SPD plasticity, (SPDP), is determined genetically. If the plasticity is 0 then the SPD is not influenced at all by the learning procedure described below, while higher values mean the SPD becomes increasingly influenced by learning.

The SPD is trained by getting the agent to produce a song according to the probabilities in its current SPD and then to 'listen' to this song with its adult SRD. For each transition in the song the algorithm looks up the corresponding probability in the adult SRD and according to this probability either 'accepts' or 'rejects' this transition. If the transition is accepted the corresponding transition in the SPD is incremented by multiplying the existing transition probability by $1 + (SPDP + \eta)$ where η is a parameter determining the learning rate (this was set to 0.1 for all simulations described here). If the transition is rejected the corresponding SPD transition is decremented in a similar way, but the change is set to $1-(SPDP+\eta)$. After the complete song has been analysed in this way the SPD matrix is normalised to ensure that the values in each row sum to 1. This process is repeated 100 times, after which the agent's SPD is said to have 'crystallised' and will not change again in the agent's lifetime.

3.2.3 Empirical plausibility

This model of song biases is admittedly simplistic as element transitions or sequence are not the only cues birds use to identify conspecific song. However, note transition matrices provides a simple and computationally tractable model of these sorts of biases. It is also in accord with one of the few empirical investigations into the neural mechanism of vocal learning in a songbird species, the white-crowned sparrow (Rose et al., 2004). White-crowned sparrows have around five acoustically distinct song elements (termed phrases in this study), which the researchers labeled *ABCDE*. Adult birds' songs are composed of these elements in a fairly fixed order. In the wild, infant birds will hear complete songs as models, but in order to investigate what cues the infant uses to memorise song the researchers tutored an experimental group of infants with phrase pairs found in normal song, but in isolation, e.g. the infants were exposed to the pair *A*-*B*, then later *B*-*C* etc. The researchers found that birds were still able to acquire a full version of song by combining the element pairs, demonstrating that exposure to the song is not necessary to acquire normal song. In another experimental group the researchers even reversed the order of the elements, training the birds on pairs from the sequence *EDCBA* and, while the group did not learn this sequence as accurately as the normal sequencing, still a number of individuals crystallised songs with the majority of phrase pairs in the order presented. The researchers argue that the lower learning accuracy of this second experiment demonstrates the difficulty for infants to completely override their genetic predisposition to normal song.

The model of song acquisition and memory suggested by this study seems consistent with the model of song biases as element transition biases described above. In the proposed model we could explain the difficulty of the second group in acquiring the reversed song as resulting from the mismatch between the genetic and learned SRD matrices, or possibly in strongly biased SPD genes.

Other recent work has investigated the neural mechanism by which birds may integrate their auditory and motor representations of song during development, i.e. how the reinforcement learning phase described above may actually be implemented in the brain. Prather et al. (2008) show that certain motor neurons in the swamp sparrow brain respond to particular note sequences when played back recordings of the bird's own song and when similar sequences are sung by other conspecifics. These neurons also respond in a very similar way when the bird sings this sequence itself and are therefore very reminiscent of the mirror neurons found in the premotor cortex of the monkey brain (Rizzolatti and Craighero, 2004). Prather et al. (2008) suggest that these neurons may allow infant birds to compare their motor actions, in the responses of these neurons, to auditory feedback systems when they are learning their adult song. The fact that these neurons respond to note sequences rather than individual notes or even entire songs provides further justification for the model of song learning as the learning of note transition probabilities. This study also provides a plausible neural mechanism for the feedback system between the developing SPD and memorised SRD that the model includes.

3.3 An investigation into some ecological conditions favouring cultural transmission

The previous section described a simple computational implementation of the auditory template model. This section describes a computational agent-based model that uses this learning model to investigate some ecological conditions that may play a role in how likely a vocal learning system is to evolve. The model is intended to investigate the conditions under which natural selection will favour increased plasticity in a population's SRDs and SPDs. For this model then, we require a model of natural selection. In order for natural selection to act on the genes encoding the SRDs and SPDs and the plasticity values we require a definition of the fitness of an agent.

3.3.1 Defining fitness

I define an agent's fitness as its ability to recognise and be recognised by conspecifics. This seems a reasonable model of one of the main pressures acting on song (Catchpole and Slater, 1995), although there are of course many other pressures acting on song in the wild as described in the previous chapter.

To calculate an agent's fitness I perform f fitness trials, set to 250 for all results shown here. In each trial the agent produces a song using its crystallised SPD. I then randomly select another member of the population and check that this second agent correctly recognises the song. An agent is said to 'recognise' a song by calculating the preference value defined in equation 3.1 of its adult SRD matrix for the song. The result of this calculation is then used as the probability that an agent recognises the song; a preference value near 1 means that the agent is very likely to recognise this song, and lower preference values mean the the agent is less likely to recognise a song. I also get the second agent to produce a song and check that the first agent correctly recognises the song. Every correct recognition means that the agent's fitness is incremented by 1.

Defining fitness in this way means that there is a strong selection pressure for the agents to develop and maintain a stereotypical and easily recognised species-specific song. As the SRD is modelled as note transition probability matrix, this corresponds to a matrix with a single high probability transition for each individual note. In short, in this environment it is adaptive to have strongly biased matrices.

3.3.2 Modelling natural selection

In order to model the natural selection of agents' genes according to the above definition of fitness I use a standard Genetic Algorithm (GA) (Mitchell, 1998), a well-known computational technique for modelling evolution. Once each agent's fitness has been calculated the GA acts on the population's genes to derive the genetic makeup of the next generation. To model selection I use a standard technique known as 'tournament selection' with a tournament size of 4. Under this regime two 'parent' agents are selected from the population who will pass on their genes to the next generation. Each parent is selected by picking 4 agents at random and picking the agent with the greatest fitness. This regime ensures that the genes of fitter individuals are more likely to be passed into the next generation. To model genetic recombination, the chromosomes of the two parents are then subject to 'crossover' with a a fixed probability p_{CO} (the crossover rate, set to 0.7 for all simulations described here). If this pair are to crossover then a random point on one of their chromosomes is selected and two 'child' genotypes are created by concatenating the portion of the first parents chromosome up to the crossover point with the remaining chromosome of the second parent, and vice versa for a second child. If the pair do not crossover their genotypes are passed on unchanged as the child genotypes.

To model genetic mutation, each gene in the two child chromosomes is subject to a low probability of mutation μ_g (the mutation rate, set to 0.01 for all simulations described here). If a gene is to be mutated it is simply replaced with a random real number in the range 0-1.

3.3.3 Modelling cultural transmission

In the model adult phenotypes are affected by the cultural environment as well as their genetic endowment. This is modelled very simply by simply recording a set number of songs produced by each adult agent and then exposing the infant agents to a randomly selected subset of this to select their model songs from, as shown in the agent lifecycle sketched below.

3.3.4 Overall model design

The overall model works with an evolving population of 100 agents. As we want to investigate how a genetically specified song can come to be learned I initialise the agents innate SPD and SRD genes to one particular song "abcd", and the plasticity genes to 0. This means that the population will start off receiving maximal fitness values and any mutations that degrade an agent's ability to sing and recognise this song should be selected against. Each agent in each generation then goes through the following 'life stages':

Birth The agent's innate SRD and SPD, along with its SRDP and SPDP, are decoded from its genes.

Development Each agent is exposed to the songs of the previous generation, and picks those which will be used as models using its innate SRD. The agent then goes through the two stages of learning described above to give them their adult SRD and crystallised SPD.

Adulthood The agents are tested in 250 fitness trials as described above to see how many times it can correctly recognise a conspecific and how many times its song is correctly recognised by a conspecific. These values are summed to give a bird's fitness score.

Reproduction Parents from the population are selected probabilistically according to their fitness score and their genes are recombined and subject to a low mutation rate to produce new child agents.

Death Each bird in the population is sampled 5 times and the resulting songs are stored for the next generation to learn from. All of the current birds in the population are removed and their children become the new population.

This process is repeated over many generations and I record various measures over the course of a run. A graphical representation of the agent's lifecycle is shown in figure 3.2.

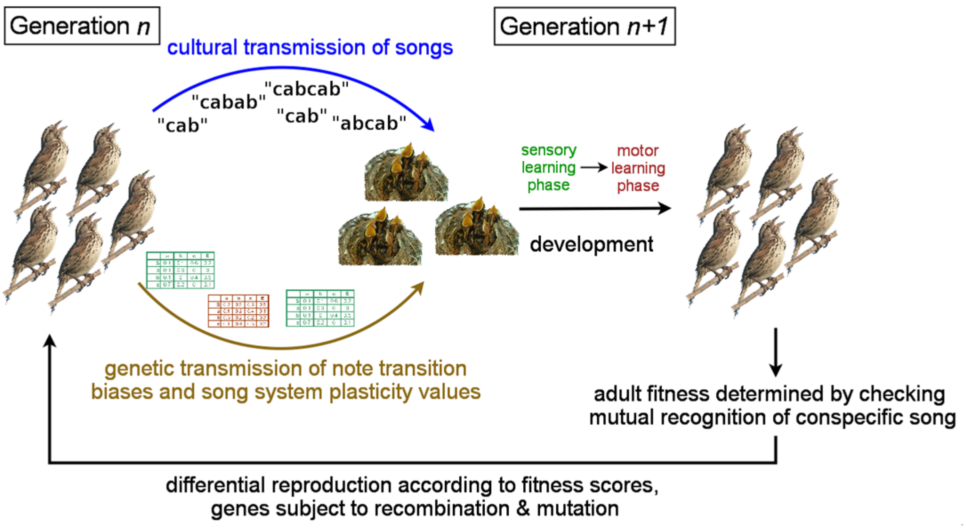


Figure 3.2: A graphical depiction of the population life cycle for this model.

3.3.5 Experiments

In this initial investigation I model two very simple ecological conditions:

Environmental reliability For the first experiment I vary the reliability of the environment, that is the degree to which the previous generation's songs are faithfully recorded and then passed on to the new generation to learn from. There are two conditions: a reliable environment where I keep 80% of the previous generation's songs, and an unreliable environment where I keep only 20% of the previous generation's songs. The remaining songs are randomly generated songs which use the same notes and are constrained to within the same length as the agents' songs. This intended to model heterospecific song or other extraneous sounds in the birds' environment.

Timing of song requirement In the first experiment I only test the bird's fitness after learning has taken place, in this second experiment we also check the bird's fitness *before* learning. This is intended to model a possible environment in which song is required immediately after birth as well as later in life.

3.3.6 Results

Results for each of the three different conditions described above are shown in figures 3.4, 3.3 and 3.5. The measures shown in each graph are the population average fitness, SPDP, SRDP, SPD change and SRD change. The SPD and SRD change are simply the absolute difference of the bias strength value of the innate and adult matrices given by equation 3.3. This measure shows us the effect that cultural transmission has on determining the adult matrices. I measure this as well as the plasticity values as the plasticity values can vary *without* a correlated variation in the change values (as demonstrated in figure 3.5). These graphs show the average values of each simulation being run 10 times.

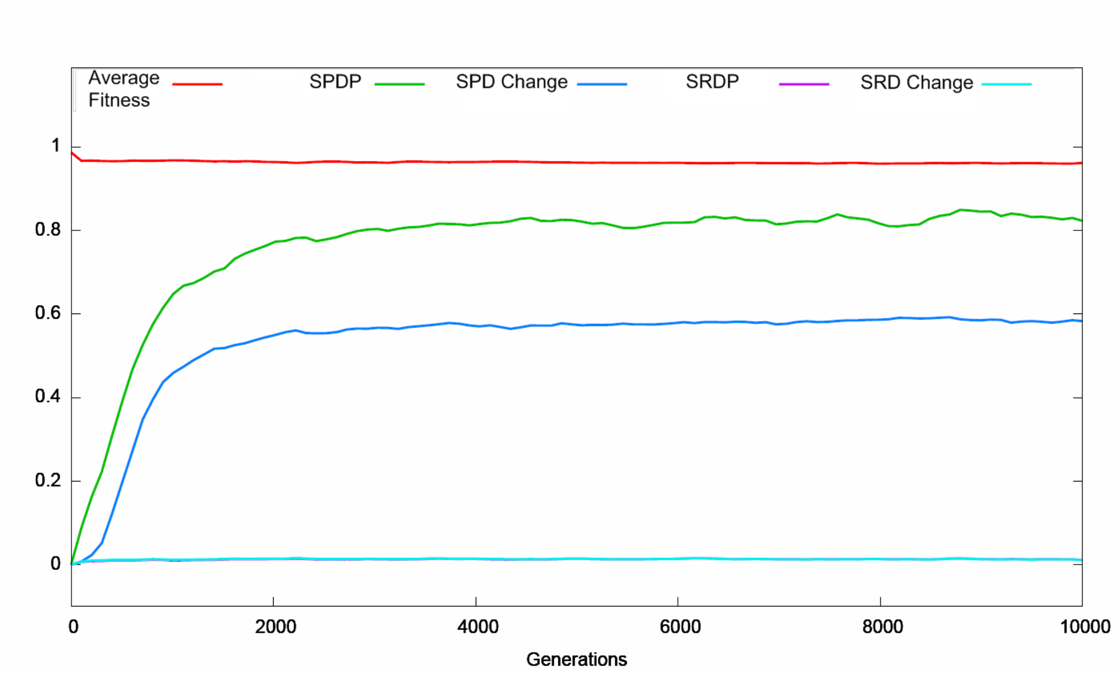


Figure 3.3: Results for the unreliable environment where only 20% of the previous generations songs are faithfully passed on. Note that in an unreliable environment the auditory template must be transmitted genetically, but motor production can be trained using this template and the genes coding for the SRD are free to drift. The SRDP value remains fixed at 0 throughout the run and so is hidden behind the SRD change line.

In all of the conditions we find that fitness stayed fairly fixed throughout all of the runs. However, the degree to which song remained being transmitted genetically depended on the environment, as demonstrated by the different values of SRDP and SRD change at the end of each simulation.

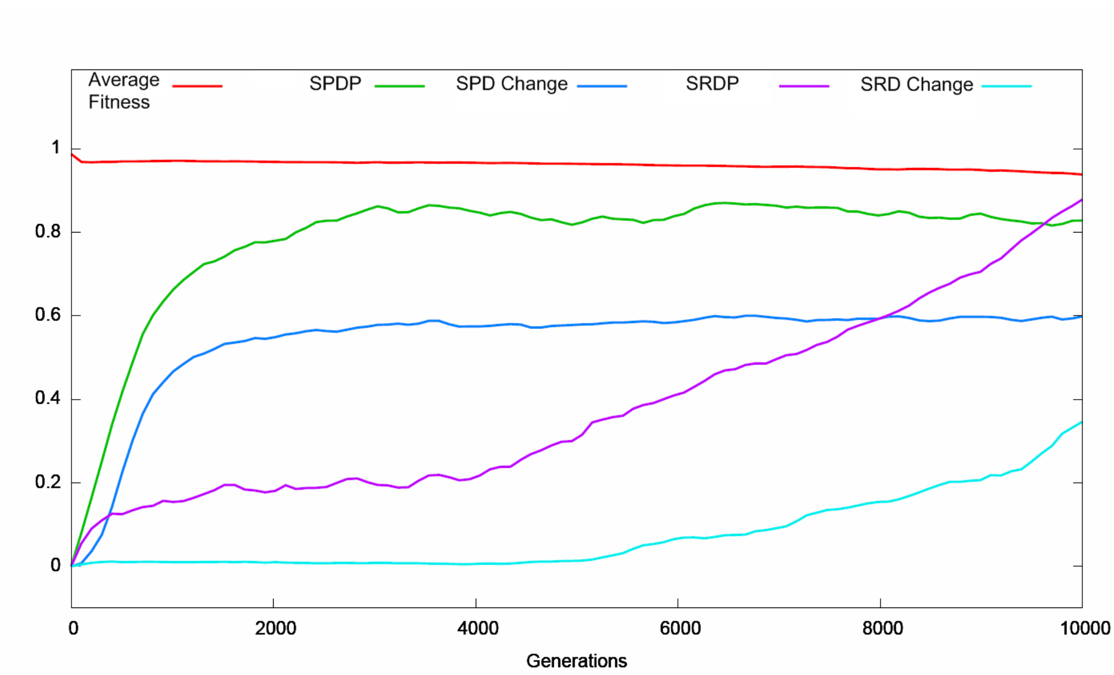


Figure 3.4: Results for the reliable environment where 80% of the previous generations songs are faithfully passed on. In a reliable environment we see that the auditory template can be transmitted culturally, and again motor production can be trained using this template.

In the unreliable environment the population cannot count on hearing conspecific song as infants. The agents therefore have to keep transmitting their song genetically, as demonstrated by the much lower SRD change and SRDP at the end of the run in figure 3.3. In contrast, in the reliable environment shown in figure 3.4, towards the end of the runs the population begins to transmit their song culturally as demonstrated by the coincident rise in the population's SRDP and SRD change.

In both experiments, however, the SPD change and SPDP quickly rise, indicating that the SPD is always being trained using the adult SRDs and the reliability of the environment appears to have no bearing on this. As long as the adults can construct a faithful copy of conspecific song in their SRDs as a result of either cultural or genetic transmission, it can always be used to train the SPD. There is no pressure for the copy of song stored in the SPD to be transmitted genetically and mutation pressure quickly erodes the genetic copy.

Figure 3.5 show results when the timing of song requirement is changed, where I test

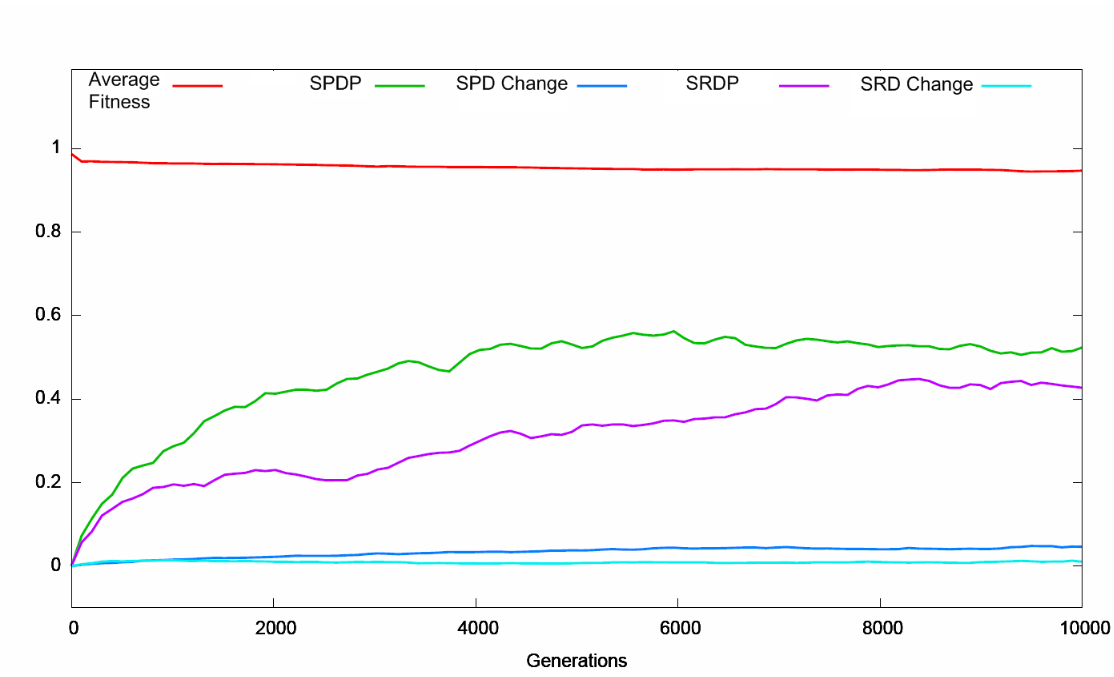


Figure 3.5: Results for the a reliable environment but where song is required immediately after birth. Under this condition both the auditory and motor copies of song must be transmitted genetically.

an agent's fitness both before and after learning. The SPD and SRD change values stay low throughout the run, demonstrating that SPD and SRD copies of song remain genetically transmitted throughout the run. The average SPDP and SRDP values drift to around 0.5 as there is no selection pressure acting to maintain these at any particular value.

3.4 Discussion

The results described here predict two simple environmental conditions which could affect the transition to a culturally transmitted communication system; the reliability of the cultural environment, and stage of life at which communicative behaviour is required. These conditions are independent of any particular assumption about the function of song learning.

I think that this model also provides an interesting case study of the interaction of genetic and cultural transmission and phenotypic plasticity. We see that where the environment is reliable enough, and 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. Cultural transmission can thus, in some conditions, be seen as a *masking* force Deacon (2003) on genetic transmission, with a similar end result to that we will see in the next chapter, for rather different environmental conditions.

Another interesting result is that in all of the experiments described here the agents come to rely solely on their auditory copy of song (in the SRD) to guide later production behaviour. I feel that this again represents a form of genetic parsimony, as it seems rather inefficient for an agent to store two 'copies' of their song genetically, even though these copies are likely to be represented in rather different ways; one being a sensory and the other a motor mechanism. Nevertheless, if there is enough phenotypic plasticity to allow these to interact, and if the genetic 'cost' of this plasticity is lower than the cost of encoding a song genetically, we see that even in the unreliable environment the agents rely on only their auditory copy, but need it always be this way round? In the case of bird song it seems so, as a bird only needs to produce a song when it is sexually mature while it needs to be able recognise conspecific song earlier. This means that the song recognition system should be more genetically constrained than the song production system, a phenomenon we observe in this model which seems to match the biological data. While this may be true of bird song it is not so clear for human language as children become capable talkers well before puberty.

3.5 Conclusion

The model introduced in this chapter provides a concrete instantiation of each of the components of the widely accepted auditory template model and I argue that, while simple, the modelling techniques seem to align well with recent studies investigating the neural implementation of song learning. The experiments I perform in this chapter investigate two rather simple conditions that may affect the evolution of a vocal learning system, but as the results demonstrate, I still find some surprisingly subtle interactions between the two representations of song that have not been discussed elsewhere. In the next two chapters I investigate some further issues that may effect the transition to a culturally transmitted system, and I also begin to investigate the role that song complexity may play alongside vocal learning. I continue to use the general framework introduced in this chapter but will model the actual components of the auditory and motor systems in somewhat different ways as the different models require. In particular, because the results presented here demonstrate that individuals need not maintain two genetic 'copies' of song as long as song is only required after sexual maturity, the models presented in the next two chapters will only consider genetic biases in the auditory system, because song production can always be trained using auditory feedback.

Chapter 4

A role for the masking of selection pressure

In this chapter I adapt some of the modelling techniques from the previous chapter to investigate a biological case study of the evolution of complex 'syntactic' structure and increased influence from cultural transmission in the Bengalese finch (*Lonchura striata* var. *domestica*). The song of this sub-species is of particular interest as it is a domesticated strain of a wild species, the white-backed munia (*Lonchura striata*), which sings a strikingly simpler and apparently less plastic song. I use the model to investigate the plausibility of two competing hypotheses for this apparent increase in song complexity; sexual selection in form of female choice, and a novel hypothesis, due to Terrence Deacon, that proposes a role for the masking of natural selection due to domestication. The modelling work demonstrates that such a significant change in the fitness landscape may play a surprising role in explaining the difference in song between the two sub-species. I discuss the possibility that a similar effect may have played a role in the evolution of human speech[1].

---

[1] A paper including much of the content of this chapter has been published as (Ritchie and Kirby, 2005) and in revised form as (Ritchie and Kirby, 2007). The latter version is included in the appendix. This paper is jointly authored with Simon Kirby who first introduced me to this case study and contributed to the initial design of the model. All of the implementation, experiments and analysis provided here are my own work, but benefited from discussion with SK.

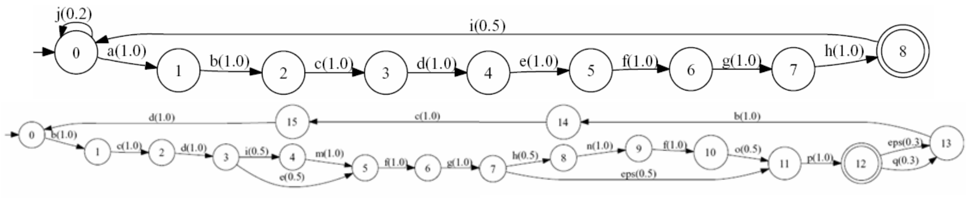


Figure 4.1: Finite state representations of a song of a white-backed munia (upper) and a Bengalese finch (lower). Note the difference in the apparent complexity of the two songs. These representations are redrawn from (Okanoya, 2002). (Okanoya, 2002).

4.1 A tale of two finches

Recent studies by Kazuo Okanoya and colleagues (e.g. Honda and Okanoya, 1999; Okanoya, 2004) of the Bengalese finch, a domesticated species of finch commonly kept as a pet in Japan, and its feral ancestor, the white-backed munia which still lives in the wild throughout Asia, provide an interesting case study of the interaction of learning and evolution in bird song. The Bengalese finch sings a song which is heavily influenced by early auditory experience and which has an apparently complex 'syntax', which Okanoya demonstrates can reasonably be described by a finite state grammar (Okanoya, 2002). In these studies Okanoya uses song linearity as a quantitative measure of song complexity, this is defined as the total number of unique song notes divided by the number of unique note-to-note transitions. While this seems a reasonably intuitive measure of song complexity I discuss some issues with it in section 4.3.2 below. Surprisingly, the munia sings a strikingly simpler, more linear song which is significantly less varied between individuals (Honda and Okanoya, 1999). It appears that even in this relatively short period of domestication, there have been radical changes in song behaviour. This has happened even though the domesticated species has (apparently) been artificially bred for plumage rather than song (Okanoya, 2002). An example finite state representation of a song from a white-backed munia and a Bengalese finch is shown in figure 4.1.

Okanoya and Hosino (2000) investigated the neural mechanism underlying this difference in behaviour and found that a bilateral lesion of the NIf song control nucleus in a Bengalese finch resulted in an individual singing a significantly more linear song than previously. A NIf lesion performed on a zebra finch, which sings a comparatively linear song, had no detectable effects. Okanoya therefore argues that the NIf, which sends auditory and motor input to the HVC, mediates song complexity in the Bengalese finch permitting the more complex phrase to phrase transitions.

Okanoya et al. (2005) performed a series of experiments investigating the female response to the more complex song of the Bengalese finch in females of both sub-species. In this study the song of one male Bengalese finch who sang a complex song was recorded. Using a software package, this song was then then broken down into its component phrases and two experimental songs were created, one with the original complex syntax and one with a simpler more linear structure and both were played to female Bengalese finches. The results demonstrate that the females displayed more nest building behaviour, a reasonable measure of song preference, when played songs with the more complex phrase transitions than when played the simpler songs. This study also investigated female preference in the white-backed munia. For this experiment the song of a male white-backed munia was recorded (to control for effects due to the different phonological structure of the elements found in the two sub-species). This song was then analysed in a similar manner but the elements were then artificially composed using a Bengalese-like syntax and female preference was compared for this artificial song compared to an unadulterated song from a male munia. Again the females demonstrated a preference for the song with more complex syntax.

Ongoing cross-fostering experiments (Okanoya, unpublished data), where infants of both sub-species were cross-fostered with parents of the other type, indicate that infant Bengalese finches are able to acquire reasonable models of munia song, but infant munias were not able to acquire songs with the more complex structure of the Bengalese finch.

4.1.1 Okanoya's hypothesis

As a result of the studies showing female preference for the more complex song in both sub-species, Okanoya (2002) argues that the difference in song behaviour is the result of sexual selection. He argues that, in the wild, perhaps the more varied structure of the Bengalese finch song would attract more predators, or require extra resources of some form, and these sorts of pressures held song complexity in check despite the apparent female preference for the more complex song. Domestication relieved these pressures acting on the ancestors of the Bengalese finch and song behaviour evolved to fit the female preference.

In order to investigate this hypothesis Sasahara and Ikegami (2004) developed a computational model inspired by the finch data that demonstrates that, under some assumptions about female preferences and perception, song complexity could indeed increase as a result of sexual selection. In their model both males and females are modelled as having individual finite state machines (FSMs) which model song production and preference. The males use their FSMs to produce song composed of distinct notes (represented as single characters, as in the previous chapter) and a special 'blank' note sequenced according to the transitions in their FSM. The blank note is intended to model the end of a song. The females assess a male's attractiveness by 'listening' to a song and updating the internal state of her FSM according to the transitions in a male's song. If a female finds herself in an 'accepting' state (that is, the song she hears is in the (regular) language recognised by her FSM) she interjects. If this interjection coincides with the song finishing (when the male produces a blank) this interjection is described as successful. The communicative success of a pair is calculated according to the number of successful interjections *except* if the female did not make at least one 'mistake' in interjecting, in which case success is defined to be 0. This is intended to model a female bias for novelty. Once all the females have listened to a male the next generation of agents is formed by probabilistically selecting parent agents according to their communicative success and passing on their FSMs to the infant agents. The FSMs are also subject to 'mutation' which can change the direction of some transitions, the number of nodes, or create an entirely new random FSM. Sasahara and Ikegami (2004) performed a number of experiments with this model and show that even if the population is initialised with simple FSMs with only two nodes, the complexity of both the male and female song grammars, defined as the number of nodes in an FSM divided by the number of transitions, does increase over time as we see in the Bengalese finch.

This model does demonstrate that an increase in song complexity could arise as a result of a female preference for novelty (though I suspect that the random mutation of FSMs may also contribute to this effect). The model does not, however, explain what held song complexity in check in the munia. This study also does not include any model of cultural transmission of song, which is a dynamic that I would argue must have played some role, or at least must be considered, in the evolution of the more complex song of the Bengalese finch.

4.1.2 Deacon's alternative hypothesis

Reviewing Okanoya's data, Terrence Deacon (in a recent presentation and in personal communications) agrees that domestication masked the natural selection pressure keeping the munia's song simple, but argues that the increase in complexity could have happened *without* direct selection for such complexity. He argues that domestication globally shielded song behaviour from selection which allowed random genetic drift to erode innate song biases, such as biases towards particular element transitions. This allowed previously minor influences, such as mnemonic biases and early auditory experience, to have more of an effect on song structure and learning. Deacon also argues that this masking of selection may also result in the various neural modules involved in song production and learning becoming increasingly *de-differentiated*, with new nuclei, such as the NIf region identified by Okanoya, playing more of a role in song behaviour. Deacon (2003) argues in detail that this process of masking of selection and subsequent de-differentiation is a potential explanation for the evolution of complex functional synergies such as the neural mechanisms for song production now present in the Bengalese finch, and, he argues, in the human capacity for language.

Deacon (2003) suggests a possible example of this sort of phenomenon; the loss of endogenous ascorbic acid (vitamin C) synthesis in anthropoid primates. He proposes that the ready availability of vitamin C in the fruit rich diet of our ancestors masked selection maintaining our ability to synthesise this vital nutrient. Mutations affecting the gene(s) involved in this process were not selected against leaving us now effectively trapped in a fruit-eating lifestyle. Wiles et al. (2005) developed a computational model of this process and show that such a masking process can indeed occur *in silico*. Their model also investigates Deacon's notion of unmasking, as if the ready supply for vitamin C in the environment was later reduced, they show that selection will favour any abilities that increased our abilities to find vitamin C rich food, such as colour vision and taste preferences for fruit. This later unmasking process will therefore bind together these varied abilities, presumably supported by genes at many more loci than the original vitamin C synthesis process. These sort of changes in the fitness landscape of a species may therefore provide a mechanism for selection to favour complex synergies such as colour vision.

4.2 A computational model

In order to evaluate Deacon's hypothesis and to try to establish if such behaviour could evolve spontaneously as a result of domestication, I have developed a computational model of the finch data related to the model presented in the previous chapter, but with some alterations to better match the empirical observations of this case study. The model is designed to be reasonably biologically plausible, and also general enough that it could be extended to other species. The model is again a computational, agent-based model which works with an evolving population of agents, or birds, and details of each stage in the simulation are given below.

4.2.1 The song filter

As in the model from the previous chapter, a bird is modelled as having a genetically coded note transition matrix, which specifies a transition probability from each note to every other note used in the simulation. The total number of notes is a parameter of the simulation, n_{notes}, but in all results provided here this was set to 8, i.e. the notes from a alphabetically through to h. This value was chosen as it appears to be around the number of unique notes identifiable in both the Bengalese finch and munia's songs (Okanoya, 2002).

As with the SRDs in the previous chapter, the transition matrix serves one main purpose; to establish the probability that a given song is one of the bird's own species song. However, in this model the matrix may be considered as a form of song 'filter'. A song with a high probability will be more likely to pass though the filter than one with a lower probability, in the example matrix in table 3.1 *cab* would be much more likely to pass through the filter than *bac*. If the matrix has a strength value as calculated by equation 3.3 near 0, i.e. it has a single high probability transition for each note this can be thought of as a strong filter, as it will only accept songs which contain these transitions. If the strength value is near 1, the matrix has more even probabilities for each transition and it is considered a weak filter as it accepts all songs equally.

In the model a bird uses its filter for two purposes:

1. To select its training set (the songs it will later use to learn from) from the songs it is exposed to during infancy.

2. As with the previous model, to judge whether another bird is a member of the same species for mating or territorial defense.

In this respect, this model is similar to those used in Lachlan's models of the 'cultural trap hypothesis' as discussed in section 2.3.4 and depicted in figure 2.3. This filter is therefore a model for an infant bird's innate preference for conspecific song.

4.2.2 Selecting the training set

An infant bird is again exposed to a set of m (set to 100 for all results described here) environmental songs randomly selected from the songs sampled from the previous generation and selects t songs as model songs. t was set to 5 for all results shown, this seems a rather low number of model songs, but the song production learning algorithm I use in this model is very computationally intensive and so a low value is used to speed up the simulation.

The infant bird is exposed to each of the m songs in turn and uses its filter and the preference value calculated using equation 3.1 to probabilistically select t model songs. The model songs selected are then fed into the learning algorithm described below.

4.2.3 Song learning

As discussed above, Okanoya (2002) argues that the songs of both the Bengalese finch and the munia can be reasonably modelled as having finite state syntax. I therefore assume that the agents in the model are capable of inducing a finite state machine representation of their song. This is a more powerful representation of song structure than can be captured by the note transition matrices (which essentially represent note bigrams) that I used in the previous chapter. For this model I therefore follow Sasahara and Ikegami's (2004) approach and represent the song production device as a (probabilistic) finite state machine.

An infant bird is assumed to use the training songs selected as described above to induce a finite state machine that represents its adult SPD. For any given set of songs there is likely to be a large number of FSMs that could have produced this data. An elegant approach to picking a likely FSM given some training data is to select the FSM with the minimum description length (MDL). The MDL principle can be seen as a formalisation of the principle of Occam's Razor. It states that the best hypothesis (candidate FSM) for some observed data (the training songs) is the one that minimises both the encoding length of the hypothesis and the encoding length of the data (Grunwald, 2007). I use Teal and Taylor's (2000) definition of the MDL of an FSM which I briefly describe here.

An FSM is formally defined as a five tuple $\{Q, \Sigma, \delta, q_0, F\}$ where Q is a finite set of states, Σ is a finite set of symbols known as the alphabet, δ is a transition function which describes all possible transitions between states $Q \times \Sigma \rightarrow Q$, $q_0 \in Q$ is the initial state and $F \subseteq Q$ is the set of final states (Sipser, 1997). Given an FSM (Teal and Taylor, 2000) define the MDL of an FSM as the sum of the data encoding length DEL, which is the number of bits required to describe the data in terms of a given FSM, and the grammar encoding length GEL which is the number of bits required to describe the FSM itself. These values are defined in equations 4.1 and 4.2 below. In equation 4.1 t is the number of songs in the training set , $|s_i|$ is the length of the ith song s_i and $z_{i,j}$ is the number of transitions out of the state reached on the jth note of song s_i. In equation 4.2 $|\delta|$ is the number of state-symbol-state triples in δ, i.e. the total number of transitions in the FSM, $|Q|$ is the number of states, $|\Sigma|$ is the size of the alphabet and

$|F|$ is the number of final states in F.

$$DEL = \sum_{i=1}^{t} \sum_{j=1}^{|s_i|} log_2 z_{i,j} \tag{4.1}$$

$$GEL = |\delta| \cdot [2 \cdot (log_2|Q|) + log_2|\Sigma|] + |F| \cdot [log_2|Q|] \tag{4.2}$$

In the model $\Sigma = \{a, b, c, d, e, f, g, h\}$ and represents each of the possible notes a bird can sing. The learning algorithm works by firstly establishing the maximal FSM that explicitly represents each song in the training set, the prefix tree. The algorithm then searches through every pair of nodes in the FSM searching for nodes which can be merged which will reduce the MDL of the overall machine, whilst also ensuring that the FSM remains deterministic (i.e. that there is no more than one transition from each state for any one symbol). When two candidate states are being considered for merging, all transitions which point *to* either of the two candidate states are adjusted to point to a new merged state, and all transitions *from* each of the candidate states are adjusted to leave the new state. The MDL of the new machine is then established according to equations 4.1 and 4.2. If the MDL of the FSM with the new merged state is lower, then the new FSM replaces the old one and the search for possible merges starts again. This process continues until no merge reduces the MDL any further. This algorithm is not guaranteed to find the FSM with the optimal MDL, but it often can significantly improve on the MDL of the prefix tree.

Essentially the algorithm searches for the most parsimonious machine in terms of the data. This approach allows a bird to generalise from its training set, whilst also always being able to reproduce each of the songs it learned from. The only difference between Teal and Taylor's and the approach used here is that I also take into account the probability of each note transition, given the probabilities of each note transition in the training set, this results in a probabilistic FSM (PFSM). When a bird sings it starts at q_0 and follows the transitions according to the probabilities associated with each transition, singing the note associated with the transition. When the bird reaches a final state it stops singing. Some example PFSMs are shown in figures 4.4 and 4.5

below, the start state is marked as state 0 and a final state is identified with a double circle.

4.2.4 Calculating a bird's fitness

To establish a bird's fitness I use the same simple model of song function as in the previous chapter; I check that a bird is able to correctly identify a conspecific using its SRD and that the bird's song, as produced with its PFSM, is correctly identified by another conspecific. Note that we do not include any model of preference for novelty or complexity.

To calculate an individual bird b_i's fitness we perform f fitness trials, a parameter set to 250 for the results provided here. In each fitness trial we get b_i to produce a song and we then randomly select another member of the population, b_j and check that b_j correctly recognises the song using its filter. We also get b_j to produce a song and check that b_i correctly recognises the song with its filter. Every correct recognition means that b_i's fitness is incremented by 1. With f set to 250, this means that the maximum fitness achievable is 500, or generally $2f$.

4.2.5 Agent life-cycle

The model works with a population of 100 agents in discrete generations. Each agent in each generation then goes through the following life cycle:

Birth The bird's song filter is built up from its genotype as described in section 4.2.1.

Development The bird is exposed to e songs from its environment, and, using its filter, selects t songs from which it will learn (its training set) as described in section 4.2.2. The bird then uses the learning algorithm described in section 4.2.3 to learn the song grammar it will use to sing throughout its life.

Adulthood The bird is tested in f fitness trials, as described in section 4.2.4 to see how many times, using its filter, it can correctly recognise a bird of its own species and how many times it is correctly recognised by a bird of its own species. These values are summed to give a bird's fitness score.

Reproduction Parent birds are selected probabilistically according to their fitness score and their chromosomes are crossed over using one-point crossover with probability p_{CO} (set to 0.7 for all results provided here), to give a new child. Individual genes are mutated with probability μ_g (set to 0.01 for all results provided here). As with the previous mode, if a gene is to be mutated it is simply replaced with a random real number in the range 0-1.

Death Each bird in the population is sampled s times and the resulting songs are stored for the next generation to learn from. The notes in these songs are subject to a low cultural mutation rate μ_c. If a note is to be mutated it is either deleted from the song, replaced with a random note or a random note is added to the song immediately after the current note. I provide results for two values of this parameter (0.01 and 0.05) because, as we will see, it has a significant effect on the model. All of the current birds in the population are removed and their children become the new population.

4.2.6 Modelling the finch data

This is a fairly general model and we need to set it up to match the data available on the Bengalese finch and the munia as closely as possible. The simulation passes through 2 main phases, each of which runs for 1000 generations. The phases are described below.

Phase 1 We know that the white-backed munia has a stereotyped linear song and that it seems to only be able to learn songs that match its species-specific song fairly closely (a munia cross-fostered with Bengalese parents is not able to learn all aspects of its tutor's song). In this model this corresponds to the munia having a strong filter biased to a linear song. To simulate this state I seed the environmental songs with a single song type, e.g. $abcdef$. I then run the simulation for 1000 generations using the fitness function and learning algorithm described above. To increase its fitness the population will have to evolve strong filters for this stereotypical song.

Phase 2 At the end of phase 1 we have a population of birds who sing a stereotypical song and produce offspring with a strong genetic bias to learn that song. We now require a model of domestication. I model this very simply by stopping the calculation of fitness values in the manner described above. This means that the genetic algorithm will no longer select parents according to their ability to recognise conspecifics, but will select parents at random and so natural selection will not act on the population.

The is clearly a very simple model of domestication but the definition of fitness described above is only intended to model the component of an individual's fitness related to species recognition by song, both for identifying potential mates and for territory defense. It seems a reasonable assumption that this pressure would be significantly reduced in a domesticated population, as the birds are kept in aviaries and mating is now controlled by human breeders. Of course natural selection would continue to act on the population for various other features and behaviours, but this model is only intended to investigate the possible effect of domestication on this one specific behaviour. Domestication, under this definition, can therefore be seen to *mask* selection for this specific recognition function.

Experimental results of this setup are described in the next section.

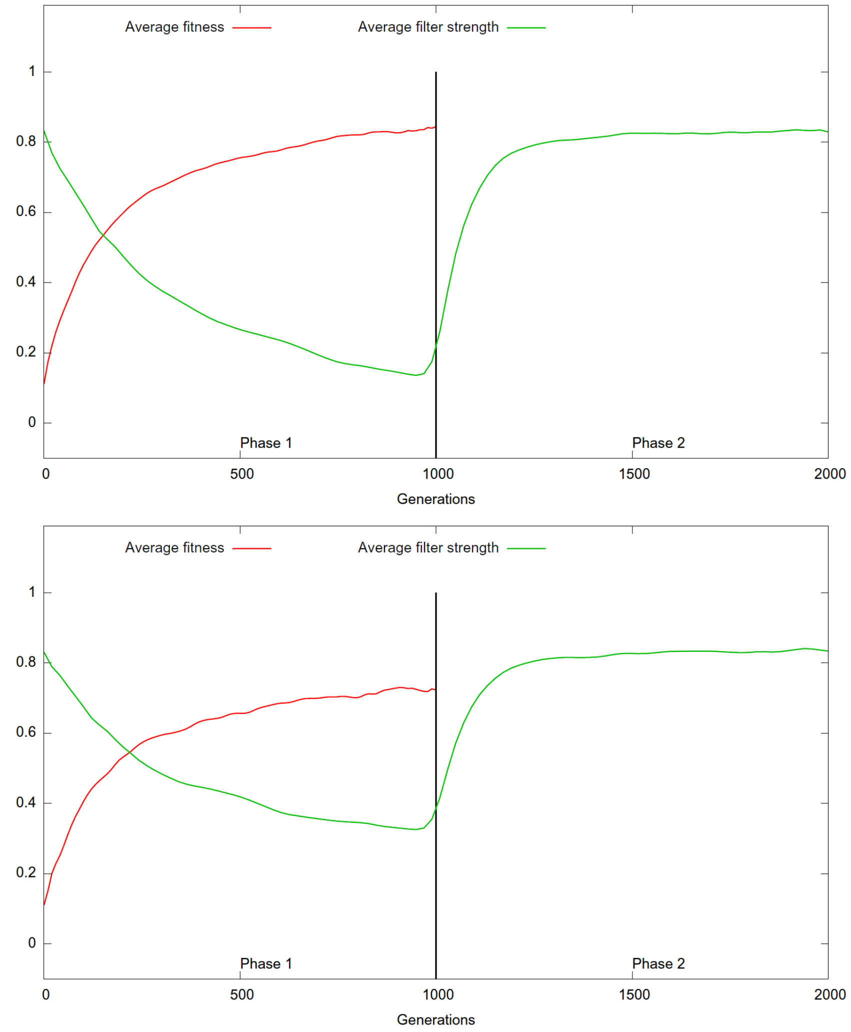


Figure 4.2: Graphs showing the change in average fitness and filter strength over the 2 phases for two values of μ_c, 0.01 in the upper graph and 0.05 in the lower graph. The graphs are scaled such that average fitness ranges between 0 and 500 (the maximum). Filter strength ranges from 0.5 to 1.

4.3 Results

The graphs in figures 4.2 and 4.3 show several measures taken over the course of each of the two phases described above for two values of μ_c, 0.01 and 0.05. These graphs are the averages of 10 runs for each condition.

The first two graphs shown in figure 4.2 show the change in the average population fitness plotted against the change in filter strengths through the two phases. Fitness

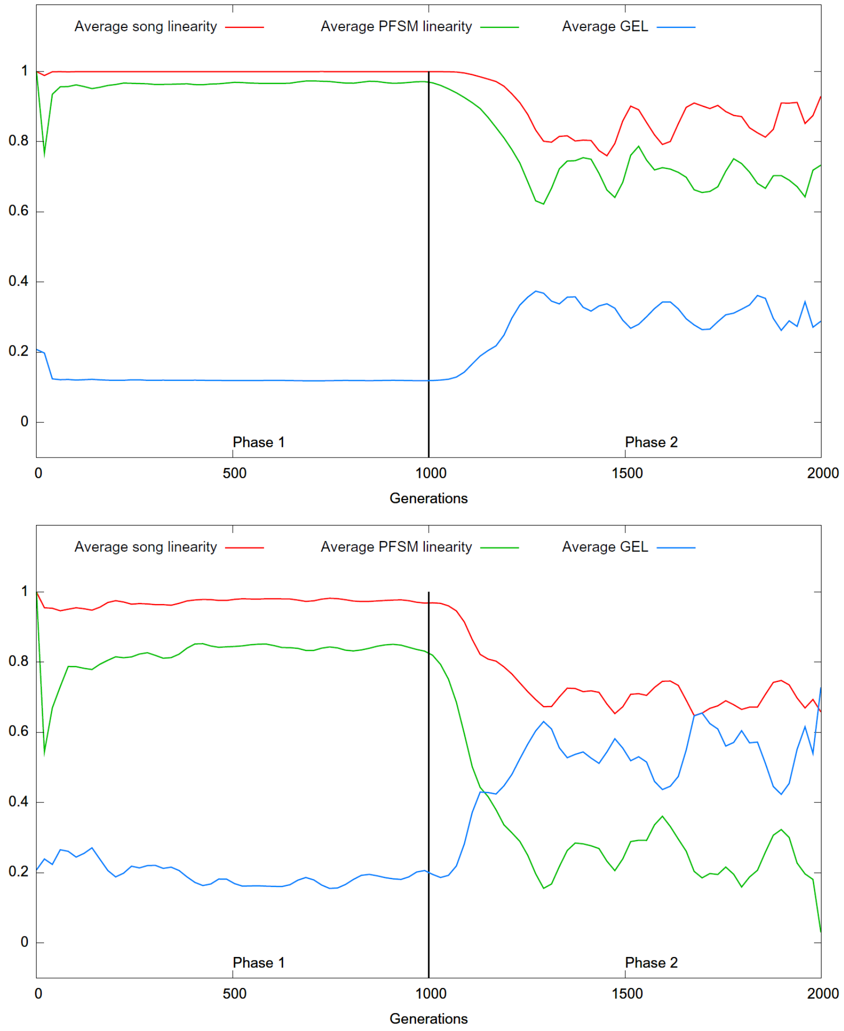


Figure 4.3: Graph showing the change in the grammar encoding length (GEL), song linearity and PFSM linearity over the course of the 2 phases for the same two values of μ_c , 0.01 in the upper graph and 0.05 in the lower graph. The graphs are scaled such the the GEL ranges from 0 to 250, and the two linearity measures range from 0.5 to 1.

values are not calculated for the population in phase 2. The filter strength is calculated as described in equation 3.3. Note that in both graphs, as the filters get stronger average fitness increases.

The second two graphs in figure 4.3 shows various complexity measures plotted against each other over the entire run, the first of these the average grammar encoding length GEL, as calculated by equation 4.2 of the population's PFSMs. The second measure is the average song linearity of the population's songs (Okanoya's definition of complexity), defined as the number of unique notes in each song divided by the number of unique note to note transitions. The final measure is the average linearity of the population's PFSMs calculated simply as the number of states divided by the number of transitions (Sasahara's definition of complexity). A completely linear PFSM with only one transition from each state would thus have a linearity of 1, while a maximally non-linear PFSM would have a linearity equal to 1 over the number of transitions in the PFSM. Note that in both graphs, all three of these values remain fairly stable throughout phase 1, but the GEL increases and the song and PFSM linearities decrease, demonstrating an apparent increase in song complexity over phase 2. In the upper graph where $\mu_c = 0.01$ we see less of an increase in GEL and more linear song in phase 2 than in the lower graph where the cultural variation is higher.

Two example PFSMs taken from the population at the end of phase 1 are shown in figure 4.4, and two PFSMs from the end of phase 2 are shown in figure 4.5. The GEL and PFSM linearity values for each machine is also given.

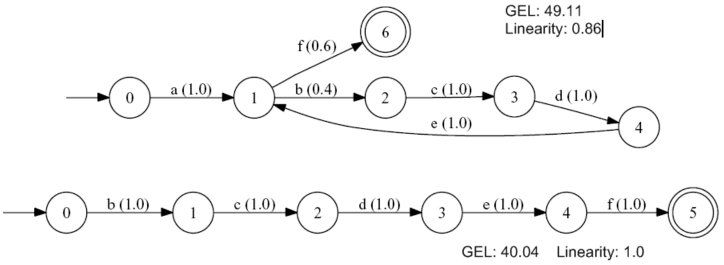


Figure 4.4: Two example PFSMs from the population at the end of phase 1.

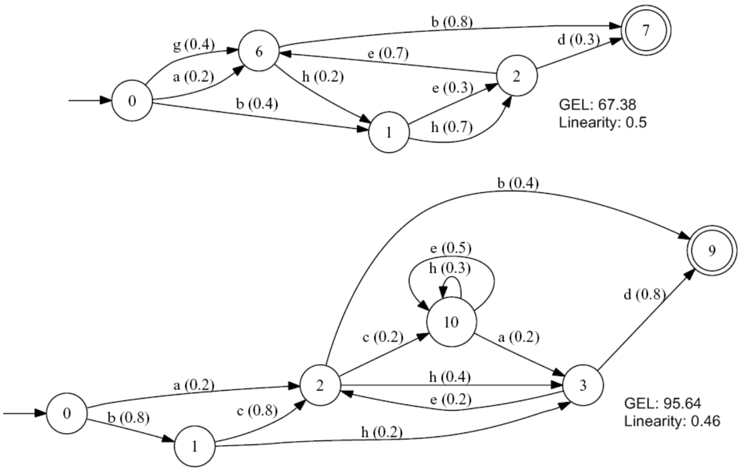


Figure 4.5: Two example PFSMs from the population at the end of phase 2.

4.3.1 Analysis

The results in figure 4.2 demonstrate that as the population's filters increase in strength the population's average fitness increases. This is because the stronger filters help the infant birds select stereotyped songs for their training set and therefore learn to produce a recognised song, and also allows them to reliably recognise a conspecific. In figure 4.3 we see that the song and PFSM linearities stay fairly high and stable throughout the phase demonstrating that the filters are selecting again any song variants that do not conform to the conspecific linear song.

When I 'domesticate' the population in phase 2 we see a significant change in its behaviour. Immediately we see that the filters begin to weaken, and we see that the average GEL also rises indicating that the population's PFSMs are getting larger and the birds have a more varied song repertoire. At the same time we see both the song, and underlying PFSM linearity drop, indicating that the songs a bird will sing have comparatively more varied note transitions, i.e. a more 'complex' song in Okanoya's sense.

This behaviour seems to be a result of the fact that the strength of the population's filters is no longer being selectively maintained, that they have been *masked* from selection. This allows mutations to accumulate and for the filters to become steadily weaker. This stops the infant birds filtering out any variant songs that may be produced by cultural mutation. This in turn results in infants inducing more varied PFSMs. Early auditory experience has come to play more of an effect on adult song behaviour. We see that where cultural variation, i.e. μ_c is higher the population appears to develop more complex song.

These results are comparable to the masking phase of Wiles et al. (2005). Their model, however, goes further than this one and shows that if the selection pressure were later *un*masked this could result in a selection pressure for any other abilities able to help the organism survive in the new environment. It would be interesting to investigate what might happen to the Bengalese finch's song if its environment (i.e. its fitness landscape) changed substantially again, e.g. a population was released back into the wild.

4.3.2 Song complexity?

Okanoya (2002) argues that the Bengalese finch has a much more ‘complex’ song than the munia. As mentioned earlier, his measure of complexity is the song linearity. He finds that the average song linearity of the munia is around 0.8 while the Bengalese finch song has a value of around 0.4. I provide results for this measure over the course of the simulations in the graph above, but on average we also see a higher value, around 0.95, for the ancestral population and a lower value, around 0.7, for the domesticated population at the end of the run.

While this measure seems a reasonably intuitive measure of song complexity (the more varied a song is, the more complex it is), it should be noted that this measure will classify an entirely random song as maximally complex. I do not want to equate randomness with complexity, but I find it hard to define a measure that can differentiate between the two. Any standard measure of the information content of a song will not be able to do so; a random song is maximally informative in information-theoretic terms. However I think that two measures, the GEL of a bird’s PFSM taken together with the linearity of the PFSM provide a reasonable approximation of the complexity of an individual’s songs. A PFSM with a very small GEL and a low linearity will produce more random songs, as it approaches a one state PFSM with multiple transitions back to the same state. An example of such a machine is shown in figure 4.6a, and has GEL of 8.0 and a linearity of 0.25 (these examples use a smaller alphabet $\Sigma = \{a, b, c, d\}$ for clarity). A PFSM with a somewhat larger GEL, but a high linearity (as we see in the ancestral population in the model) will produce a more linear song, an example is shown in figure 4.6b. A PFSM with a large GEL and a relatively low linearity will produce songs that I am more happy to refer to as complex, an example is in figure 4.6c. The larger GEL indicates that it has many states, and so different notes will be used in different contexts, but the linearity measure indicates that each state also has several transitions which means that different transitions can be made from each context. There are, though, machines with large GELs and low linearity, such as that such shown in figure 4.7, which will produce the same songs as the machine in figure 4.6a (although the song length will always be a multiple of four), and which I would therefore characterise as random. The MDL algorithm should not produce such machines though, as clearly the more parsimonious machine is the single state FSM.

However, the algorithm I employ is not guaranteed to find the machine with minimal MDL as it is only a hill-climbing search (induction of a finite state machine from given data is known to be NP-hard Gold, 1978) and so given a pathological data set it may produce machines somewhat like figure 4.7 and mislead this sort of analysis.

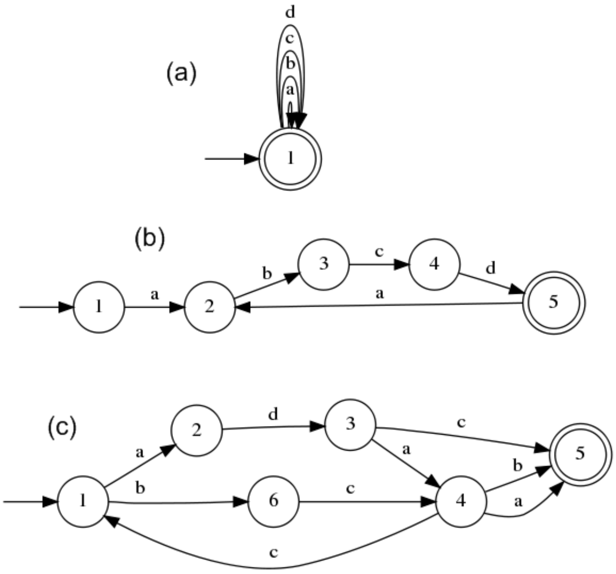


Figure 4.6: A comparison of various FSMs. (a) will produce random strings and has GEL = 8.0 and linearity = 0.25. (b) will produce linear strings and has GEL = 35.54 and linearity = 1.0. (c) will produce 'complex' strings and has GEL = 62.12 and linearity = 0.56.

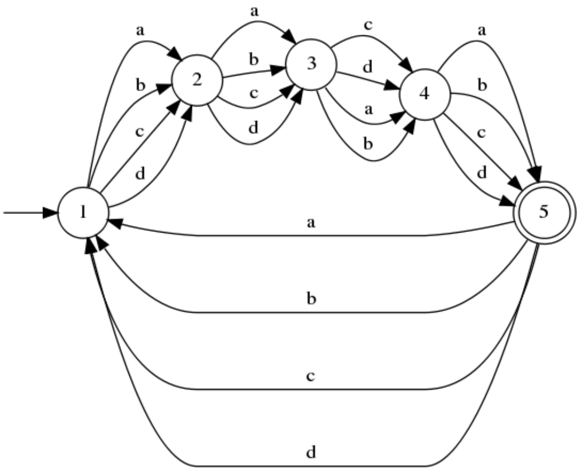


Figure 4.7: A 'pathological' FSM with large GEL (135.20) and low linearity (0.25), but which will produce strings that are as 'random' as those produced by the machine in figure 4.6a.

Nevertheless, the results demonstrate that the PFSMs of the 'domesticated' population at the end of phase 2 do have generally higher *GEL*s and lower linearity than the 'wild' population at the end of phase 1 and so I am tentatively happy to agree that domestication has caused an increase in song complexity according to the definitions employed by Okanoya and Sasahara in their studies.

It seems to me that such information-theoretic analyses of song strings is unlikely to be able to provide us with a satisfying definition of song complexity, as different researchers can use different measures to describe the same data. An observer-neutral and empirically grounded definition of song complexity would be useful in further study of the differences in song in the Bengalese finch and the munia, and indeed for other songbirds. I will propose just such a definition in the next chapter, where I argue that differences in the learnability of songs may provide us with meaningful, and testable, definition of song complexity. It would be interesting to establish if the song features preferred by female Bengalese finches are indeed harder for the males to acquire.

4.3.3 Comparison with the empirical data

Comparing these results with the data available for the Bengalese finch we find that the model does seem to capture some of the phenomena observed. The song of our 'domesticated' population appears to fit the same criteria for complexity that Okanoya proposes for the Bengalese finch; the difference in the values for the song linearity in the ancestral and domesticated populations seem to match fairly well. Okanoya has shown that while Bengalese chicks can readily learn munia songs, munia chicks cannot learn the more complex Bengalese songs. In the model this difference is attributable to the munia's stronger filters.

As it stands though, the model does not explain why the female munia prefers the Bengalese song. I would argue that it is possible a bias for complex song may have been latent in the munia, and the fact that the munia females prefer the more complex song does not prove that this was the driving force for the change in song behaviour, although as discussed above Sasahara and Ikegami's (2003) models show that this may be possible. Okanoya (2004) demonstrates that the NIf region of the Bengalese finch's

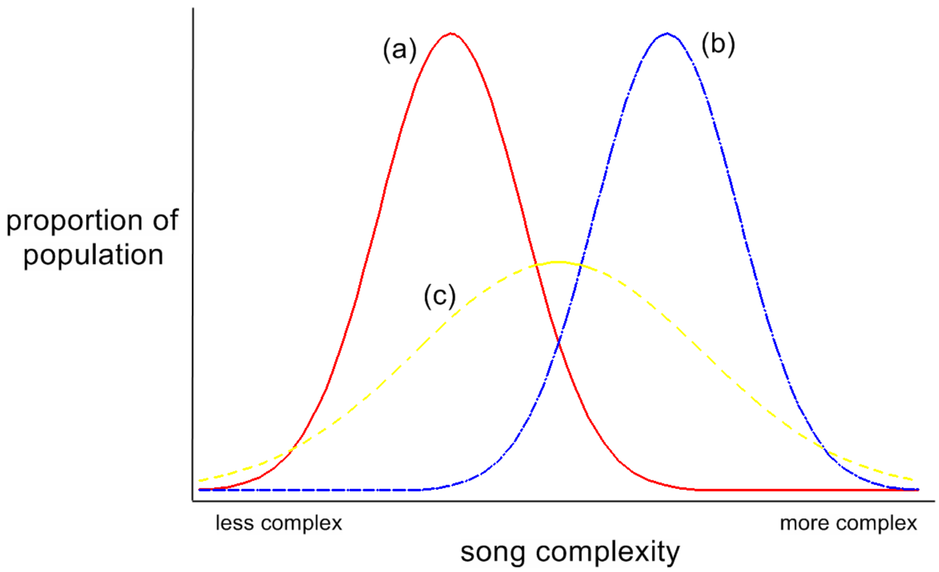


Figure 4.8: A sketch of the two possible evolutionary trajectories from the ancestral munia population represented here by (a). The x-axis represents song 'complexity' and the y-axis represents the proportion of the population with a song of a particular complexity. (b) shows the expected distribution from positive selection for song complexity, while (c) shows the expected distribution of songs if masking were more important. Note that there will likely be an increase in individuals singing more complex song just as a result of the 'slackening' of the distribution.

brain is necessary for it to be able to sing the more complex song; when surgically lesioned a Bengalese finch with previously complex song will sing a simpler, more munia-like song. I would argue that the model remains neutral to this datum, as it is possible that the munia does have this pathway present in its brain but, because it only ever learns a simpler song, does not use it.

More empirical work can be performed to investigate whether Okanoya's or Deacon's hypotheses are more likely to be correct. If masking and genetic and cultural mutation played a dominant role I suggest that we should observe more inter-individual variation in song behaviour, some individuals may sing more complex song while others may sing simpler songs. If female choice plays more of a role we should see a more uniform shift to song complexity in the population. A sketch of my predicted distributions of song complexity as a result of each mechanism is shown in figure 4.8. It is also possible that both effects played some role. Masking may be the mechanism by which the Bengalese finch acquired a more complex song, but this behaviour may then fit a latent preference for variability in the females which then imposes a new selection pressure for any mechanism that can provide such variation. The two, apparently alternative, hypotheses may thus actually just explain two stages in the evolution of this behaviour.

These arguments are supported by some recent empirical work by Okanoya and colleagues which suggests that song complexity in populations of wild munias may be affected the presence of sympatric species. These studies suggest that where there are closely related species (in this case the spotted munia) living near white-backed munias, song structure remains simpler, while where there are fewer closely related species song complexity can increase (Okanoya, unpublished data[2]). These observations fit well with the results presented here, providing further justification for the important role for species recognition as a check on the development of song complexity that we found.

4.4 Discussion

These results demonstrate that, as Deacon proposed, an increase in song complexity (in some sense) and increased vocal plasticity can arise *without* direct selection on either trait, simply through the process of domestication causing a major alteration to the fitness landscape. We see that when the natural selection pressure for stereotyped song is masked, genetic biases are quickly eroded and minor cultural variation which was previously filtered out is amplified such that song behaviour changes dramatically.

As mentioned, this model does not explain the preference for complex song in the munia or the Bengalese finch females found by (Okanoya et al., 2005). While I would argue that the process I have investigated here may have played a role in the evolution of song complexity in the Bengalese finch, if this species were once again released into the wild it would be interesting to see if whatever mechanisms that supported the new

---

[2] Presented at the “Acoustic Communication by Animals” conference, Oregon State University, 2008

song features, such as the NIf nucleus, would be maintained by the selection pressure from female preference. This would complete the cycle of masking and unmasking that Deacon (2003) proposes and Wiles et al. (2005) demonstrate.

Whether either hypothesis turns out to be right, both propose an important role for domestication for the change in selection pressures. Both Okanoya and Deacon argue that humans have undergone such an ecological change in our evolutionary past. Okanoya (2002) proposes that the development of tool making in humans in a savannah environment protected us from environmental pressures and led to a form of 'self-domestication'. He argues that in such an environment sexual selection will come to play a more significant role in shaping our evolutionary trajectory. Deacon (2003) also argues that we are a self-domesticated species, but that as well as tools, our capacity for symbolic culture masked other selection pressures and created a new ecological niche. The results presented in this chapter are in accord with these arguments and suggest that is possible that we should not be looking for selective advantages of a culturally transmitted syntactic language, but rather asking what selective forces may have been shielded or modified in our recent evolutionary past. The lifting of selection pressure, and the subsequent diversification of behaviour could have been necessary precursors of a system of iterated learning for language.

4.5 Conclusion

The model presented in this chapter builds upon the framework introduced in chapter 3 but includes a more powerful model for the song learning system that allowed us to capture the observed structure of Bengalese finch and white-backed munia song. We saw that as the fitness landscape changed when the ancestors of the Bengalese finch were domesticated, song syntax changed substantially despite there being no selection on song behaviour. This change resulted from a similar effect to that which we found in the previous chapter; when the population's environment changed mutations were allowed to accumulate and genetic information came to contribute less to the adult phenotype. This allowed more influence from cultural transmission and also lead to the song learning system inducing more varied grammars in a way that seems to capture at least some of the empirical data. As I discussed though, whether the song of the Bengalese finch is more ‘complex’ in some meaningful way than the munia is not clear, and I suggest that information-theoretic analyses such as have been previously applied and which I attempt to apply here are unlikely to provide us with a satisfying definition. In an effort to resolve this issue, in the next chapter I will explicitly link some features of song that are frequently described as ‘complex’ with fitness-relevant information and propose an empirically-grounded definition of song complexity as songs that are harder to learn. In both the models I have presented so far I have assumed a simple species-recognition function for song and song *learning* has not really been explicitly linked to this function. The model introduced in the next chapter will address this and propose a plausible evolutionary function for song learning.

Chapter 5

Song learning as an indicator mechanism

In this chapter I present several models intended to investigate a recent novel hypothesis for an evolutionary function of vocal learning in bird song, the *developmental stress hypothesis* (DSH). This hypothesis proposes that song learning may serve as an indicator mechanism allowing females to use learned features of song as a window on a male's early development, a potentially stressful period that may have long-term phenotypic effects. I first discuss the argument, setting it in the context of previous work on the sexual selection of signals, and present some empirical evidence supporting the hypothesis. I then introduce a simple but general formal framework for theoretical investigation of the hypothesis and construct several models based on this framework. The first is a simple agent-based model related to the models used in the previous two chapters and the results from this model provide some initial support for the DSH. I then develop more rigorous and general population genetic models which can provide greater insight into the dynamics of the model and the effect of the various parameters. Finally I discuss possible implications from these results for the evolution of vocal learning and perhaps other forms of linguistic and cultural learning in humans and other species[1].

---

[1] A paper including much of the content of this chapter has been published as (Ritchie et al., 2008) and is included in the appendix. This paper is co-authored with Simon Kirby and David Hawkey, and both contributed to the work described in this chapter. The general framework described resulted from

5.1 Indicator mechanisms and song learning

The models in the previous two chapters investigated some conditions under which learning can come to play more or less of a role in song transmission as a result of differing ecological conditions. We found that learning can only evolve under certain life history conditions and that domestication may play a surprising role in allowing early learning to have a greater effect on adult song behaviour. In these models learning has not really had an effect on the evolutionary function of song, except to the extent that it permits accurate species recognition. These models may therefore contribute to an explanation of the ecological conditions under which vocal learning might have arisen in the songbirds, but they do not address *why* vocal learning evolved in this group.

As we saw in chapter 2, song learning is underpinned by an intricate set of neural systems and requires a significant investment in time before a young bird will develop normal adult song. It seems that there must have been some strong selective advantages for vocal learning at some stage in the evolution of the songbirds, and for the subsequent maintenance of this system. As I discussed earlier, in section 2.3.4, there have been a number of mooted advantages to song learning, but the empirical evidence for several of these seem to be limited to one or a few species and few appear to apply widely to the many song learning species. Lachlan and his co-worker's cultural trap hypothesis (CTH) is the only hypothesis that appears to hold widely over vocal learning species (Lachlan and Slater, 1999; Lachlan and Feldman, 2003). This is an important and interesting hypothesis, but it does not propose any (traditional[2]) adaptive function for song learning; it holds that that once song learning has evolved in some form, perhaps as a result of a process such as random genetic drift, the interaction of genetic and cultural evolution, under certain assumptions, are likely to maintain the trait. While this central assertion of the CTH may turn out to be correct, given the widespread occurrence of song learning in birds, and the complex physiology that supports it, further inquiry into more traditionally adaptive functions for song learning

---

discussions with myself and SK, and DH helped with some of the mathematical analysis of the behaviour of the model, otherwise all of the discussion, modelling, results and analysis presented here are my own work.

[2] Lachlan has, however, argued (in a personal communication) that the CTH could be considered as adaptive from the point of view of the culturally evolving entities themselves.

seem reasonable.

In looking for an adaptive function of song learning, it seems sensible to look again briefly at the adaptive functions that have been proposed for song itself. As we saw in section 2.2.2.2 song is a sexually selected trait; generally males sing to attract females and repel invaders. In many species it appears that females in particular are influenced in their mate choice by several variable song parameters, such as song output and repertoire size (reviewed in Searcy and Yasukawa, 1996). Acquiring a large repertoire is one of the more accepted proximate functions of song learning, but why should a female pick a male according to the size of their song repertoire?

It may be that a feature such as a large repertoire is simply more stimulating to females, perhaps because it makes the male more noticeable in the face of general habituation processes in the female auditory system. This could then lead to a Fisherian runaway process (Andersson, 1994), whereby this existing sensory bias creates a selection pressure for males to have more variable song and any mechanism capable of supporting this, such as vocal learning, would be selected for. This effect could then be amplified as males with more variable song would then be more attractive mates. It would then be beneficial for females to bear sons with this trait as they will carry the females genes as well; thus producing ‘runaway’ selection for the trait. Several mathematical models have convincingly shown that this runaway process can work in theory, and there is some supporting empirical evidence, though it is far from conclusive (reviewed in Andersson, 1994).

In some cases though, variation in song parameters that females attend to has been shown to correlate with factors that seem clearly indicative of male quality. For example, Hasselquist (1998) demonstrates that the syllable repertoire size of a male great reed warbler is a good predictor of the post-fledgling survival of his offspring. In this case it appears that the females preferences are not arbitrary and that it is in fact rational to pick males with larger repertoires, but it is not clear *why* a larger repertoire of syllables should correlate with male quality.

A signal that correlates in some way with the condition of the signaller, as appears to be the case in the great reed warbler, is known in sexual selection theory as an *indicator*.

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. Without such a mechanism there would be a strong selection pressure for all individuals to signal that they are high quality and the signal would become useless to a female trying to pick between mates. That is, for an indicator system to be evolutionarily stable there must be some mechanism ensuring the *honesty* of the signal.

There have been a number of proposed ways in which signal honesty can be maintained but Maynard Smith and Harper (1995) identify two important mechanisms that seem applicable to bird song; *indices* and *costs*[3]. They define an *index* as a "signal that is physically associated with a quality of interest to the receiver", such as tigers (*Panthera tigris*) marking territory by scratching as high as they can on a tree trunk, thus reliably signalling their size. A *cost-added signal*, on the other hand, is defined as "a signal that is more costly to make than the minimum required to transmit the information" and was originally termed a *handicap* by Zahavi (1975). If a signal has some cost to produce - as will nearly always be the case - but also conveys some benefit to the signaller, such as attracting more mates, for each individual there will be some optimal level of signalling where the costs balance the benefits. Signal reliability may then be ensured if the costs fall on signallers according to their quality, such that lower quality individuals incur greater costs than high quality individuals. This will result in a lower optimal level of signalling for lower quality signallers. Essentially only high quality signallers are able to 'afford' to signal at a higher level. The consistency of this idea has also been convincingly demonstrated with theoretical work (e.g. Grafen, 1990). As Maynard Smith and Harper (1995) note, in a model the difference between indices and costs is clear but differentiating between these two possible mechanisms is not a trivial task when dealing with a real biological example.

An indicator trait may provide a female with information about a male in two possible ways. The trait may be linked to the genetic makeup of the male in some way enabling a female to select males with "good genes", this is known as a *indirect* benefit as it will benefit her only through passing these good genes onto her offspring. The indicator

---

[3] Maynard Smith and Harper (1995) also identify a third mechanism that they term a *minimal signal* which can be stable because the signaller and receiver share a common interest, such as communication between mated pairs or genetic relatives. This does not seem applicable to bird song as in mate choice males and females interests undoubtedly differ, it may apply, however to other bird calls.

trait may also be linked to the phenotypic quality of a male, such as his ability to defend territory or find food, in which case the female enjoys *direct* benefits. Theoretical models have shown that female preferences can evolve to use indicators when she receives both direct and indirect benefits (Andersson, 1994).

Given that a male bird's song clearly influences a female's mate choice, and that females are apparently able to use song to identify high quality males, it seems likely that song may serve an indicator of male quality. To demonstrate this, however, we must provide evidence linking a male's quality and his singing abilities. Song may serve as a costly signal whereby only high quality males are able to produce songs with certain characteristics, some possible costs of song production include:

- Energetic costs: Thomas (2002) demonstrates that male nightingales (*Luscinia megarhynchos*) may lose up to 5-10% of their body weight in a single night of singing. However, this may be an exception as Oberweger and Goller (2001) show that the metabolic costs of singing for three songbird species is no higher than the costs of various other types of vocal behaviour in other bird groups.

- Time costs: time spent singing may be better spent foraging or establishing territory.

- Increased risk of predation: singing may attract females, but it will also allow a predator to notice and locate the singer.

These sorts of costs may be able to explain the observation that females prefer males with greater song output, i.e. males who simply sing more, as a male in better condition is likely to be able to spend longer singing and more able to escape predators. However, it is not obvious what role any of these sorts of production costs could play in linking male quality with several of the other features that have been clearly demonstrated to affect female preference, such as increased song complexity or more accurate song learning. As an example, it is not clear that a bird's ability to acquire a larger range of syllables will mean that it expends more time or energy, or is more likely to be predated upon, than a bird only able to acquire a simpler song. Yet, as we saw earlier, female great reed warblers appear to prefer males singing songs with more syllable variation,

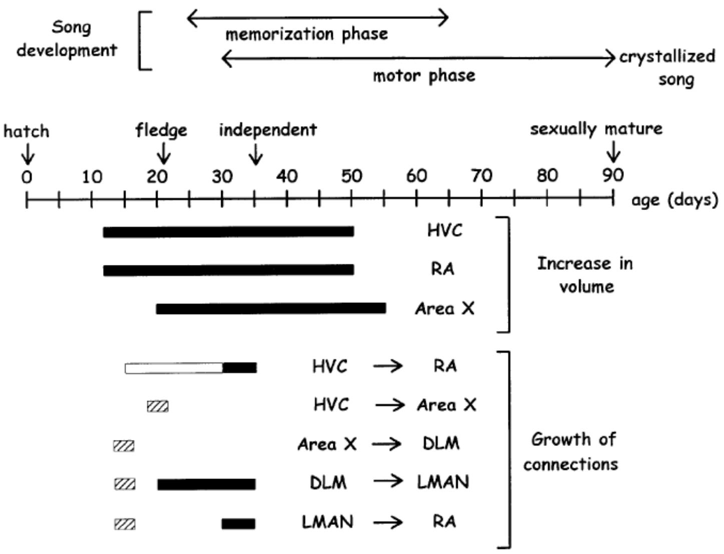


Figure 5.1: Time line of zebra finch development, showing the coincidence of song learning, significant life history events and song system development. See section 2.4 for an explanation of the various acronyms for the brain regions identified. Figure from (Nowicki et al., 1998), reproduced with permission.

and males singing in this way appear to be more successful parents.

In order to address this apparent quandary and to provide a possible link between male quality and these sorts of female preferences Nowicki et al. (1998) proposed what they initially termed the “nutritional stress hypothesis”, which was later generalised to the “developmental stress hypothesis” (Buchanan et al., 2003; Nowicki and Searcy, 2004). This hypothesis arises from two general observations of songbird life history. Firstly, songbirds are altricial and are completely dependent on parental care for a significant period after birth (reviewed in O’Connor, 1984) and this period is fraught with potential stresses. Nestling growth and survival rates are correlated with food abundance and parental investment, and starvation is common (Nowicki et al., 1998). Even short term food deprivation as a nestling has been shown to have long-term effects on adult body size, a reliable correlate with many measures of male fitness (Nowicki and Searcy, 2004). Parasites also attack young songbirds and this too can have long-term affects on the adult phenotype (Buchanan et al., 1999). Stresses caused by differential nestling social status may also have a detrimental effect (Spencer et al., 2004).

The second observation is that the brain structures mediating song learning develop in infancy, late compared to the rest of the avian nervous system and roughly coincident with this precarious period of development, as shown for the zebra finch in figure 5.1. If developmental stress has long term effects on a male's parental abilities or if response to developmental stress reflects male genetic or phenotypic quality, any mechanism allowing females to infer a potential mate's level of developmental stress would be beneficial. Because song learning occurs during the period of life when an infant bird is most susceptible to developmental stress, features of song acquired during this period may be irreparably affected by developmental stress. Such features of song may thus be used as reliable indicators of the level of developmental stress suffered. It then seems plausible that females could evolve a preference for such features as this would enable them to identify males affected by developmental stress.

5.2 Empirical evidence for the DSH

Several of the key assumptions and predictions of the DSH have been investigated in a range of empirical studies and I provide a brief review of these in this section.

One of the key assumptions of the DSH is that developmental stress has a long-term effect on adult song. A number of possible stressors have been investigated in a number of different species, several of which do seem to affect adult song in ways that seem likely to affect its function in either attracting a mate or defending territory. A brief review of relevant work includes the following:

- Undernutrition has an effect on the accuracy of song note learning in male swamp sparrows (*Melospiza georgiana*) (Nowicki et al., 2002a)

- Social rank has an effect on acquired song phrase repertoire size in the European starling (*Sturnus vulgaris*) (Spencer et al., 2004).

- Food restriction and corticosterone administration selectively affect growth of the HVC (an important song control nucleus) in zebra finches (*Taeniopygia guttata*) (Buchanan et al., 2004).

- The song output of European starlings subject to unpredictable short-term food deprivation as nestlings differed significantly to controls in a number of ways relevant to female preference (Buchanan et al., 2003).

- Male zebra finches subject to dietary restriction develop significantly shorter song motifs and their songs use fewer syllable types (Spencer et al., 2003).

- At least one measure of nestling growth (the length of the innermost primary feather) correlates with adult repertoire size in the great reed warbler (*Acrocephalus arundinaceus*) (Nowicki et al., 2000).

Another prediction of the hypothesis is that females attend to these effects and two studies provide some support:

- Female song sparrows (*Melospiza melodia*) respond more to males that have copied the notes of their tutors more accurately (Nowicki et al., 2002b).

- Female zebra finches significantly prefer the songs of control males to those of males stressed by either food restriction or corticosterone administration (Spencer et al., 2005).

A final, important assumption is that there is a fitness detriment to mating with a male who has suffered from developmental stress. This seems like a plausible assumption, as a phenotypically weakened male is likely to be a less effective parent. Unfortunately I am not aware of any direct empirical investigation of the effect of developmental stress on male parental abilities, but Naguib et al. (2006) show that one possible form of developmental stress (in their study, brood size) of mother zebra finches has an effect on the hatching and fledgling success of their daughters one generation later. There is also some indirect evidence for the effect on male parental abilities. As mentioned before, Hasselquist (1998) shows that male syllable repertoire size is positively correlated with harem size and fledging success in great reed warblers, and Nowicki et al. (2000) show that nestling developmental stress has an effect on syllable repertoire size in this species, so it seems that developmentally stressed males may fare worse as parents.

These data suggest that several features of male song are affected by developmental stress, that this stress may have long-term fitness relevant effects, and that females attend to these features, providing support for a number of predictions of the hypothesis.

5.3 Outstanding issues

Despite this range of empirical support there remain some issues raised by the hypothesis that haven't yet been accounted for either in verbal arguments or empirical investigation and I identify some of these here.

Firstly the arguments have not generally considered the possible effect of developmental stress on female choice. Female songbirds are also altricial and may suffer from various developmental stressors in much the same way as males. Given that in many species it appears that female adult song preferences are also affected by early learning, and that females use at least some of the same neural substrate as the males, it seems plausible that their acquisition or memorisation of particular song features may be subject to a similar effect as for the males. This seems particularly relevant to species such as the song sparrow, in which females appear to prefer males who have more accurately learned the song of the local population (Nowicki et al., 2002b). In order to judge the accuracy of a male's learning the female must have acquired an accurate model of local song herself. This may then add a complication to the argument as only high quality females may be able to assert a preference for revealing song features and so the preference as well as the feature may be dependent on condition.

Another consequence of the hypothesis is that culturally transmitted song types that reveal a male's level of developmental stress cannot be acquired by all members of a population. Previous theoretical work on the evolution of cultural communication systems, especially human language, (e.g. Brighton et al., 2005; Oudeyer, 2005) has argued that there is often a cultural evolutionary pressure in such systems for the signals to become increasingly learnable. This seems an intuitive result as culturally transmitted communication systems must be able to be acquired by each new generation of learners if they are to persist, and so less learnable forms are likely to be selected against. However, the DSH assumes that song types that are not learnable by lower quality males must be able to survive cultural transmission.

In the following sections I attempt to formalise the developmental stress hypothesis in order to investigate these issues by developing a series of models of the evolution of song preferences. With the models I intend to address the two important questions raised above: firstly, can a preference for culturally transmitted features of song affected by developmental stress evolve in a population even when both males and females may be affected by such stress? Secondly, can cultural transmission of songs with such features be stable, despite the fact that those members of the population suffering developmental stress cannot acquire or produce them?

5.4 A framework for modelling the DSH

I model developmental stress simply as resulting in either a high or low quality bird, whereby a low quality bird may be thought of as undernourished or parasitised. Although quality could be modelled more realistically by being continuously distributed, this simplifying assumption makes analysis of the models more tractable. Similar simplifications have been made in previous studies, such as Heywood's (1989) 'good' and 'bad' environments used in a model of the handicap mechanism.

The probability of an individual bird of either sex being high quality is defined by a parameter δ. The factors affecting a bird's quality are assumed to be determined by the environment. Quality is therefore not inherited and is assigned to individuals independently of their genetic constitution and the quality of their parents. I assume that it would be beneficial to females to be able to identify a potential mate's quality, but that a bird's quality is not directly observable. Instead I assume, as the DSH asserts, that the quality of a bird affects its ability to acquire some song types or features of song. In the model, 'type 2' songs are only learnable by high quality birds, while 'type 1' songs are learnable by both high and low quality birds.

I make no further assumptions about the song types other than this difference in learnability, however these two classes of song are intended to model the differences in song behaviour found in high and low quality birds. It is therefore useful to think of type 2 song as being more ‘complex’ in some way than type 1 song, e.g. it may have more varied phrase or note syntactic structure, or can be thought to represent a larger syllable repertoire. Considered in this way, I am therefore assuming that high quality birds are able to learn more complex songs than low quality birds. I also assume that only high quality birds are able to differentiate between the two song types.

The proportion of type 2 songs in the cultural environment is denoted as t_2 and the proportion of type 1 songs $t_1 = 1 - t_2$. Importantly, I assume that both males and females are affected by developmental stress and that this stress can affect both sexes’ adult behaviour.

A bird’s ability to acquire a particular song type is also affected by a genetic component, which I model as a single autosomal locus A for which two alleles exist, A_1 and A_2. This gene may be considered as coding for a learning bias and, as in models from the previous chapters and in (Lachlan and Feldman, 2003), I assume that the same gene controls song learning bias in both males and females. This assumption seems plausible as studies have shown (reviewed in Riebel, 2003) that, in many species, both sexes are born with a preference for conspecific song which is affected by early song exposure (e.g. in zebra finches Lauay et al., 2004), that females share many of the song system nuclei with males, and that females can produce song under abnormal hormonal conditions (Riebel, 2003). Birds with allele A_1 can only learn type 1 songs. Birds with allele A_2 are able to learn song types 1 and 2, but are biased to acquiring type 2 song. This bias will, however, only be expressed if a bird hears type 2 song as an infant. The proportion of the population with allele A_2 is denoted as a_2, with the proportion of birds with allele A_1, $a_1 = 1 - a_2$.

This gene is intended to play a similar role to the genes that coded for the SRDs and song filters in the previous two chapters in that it represents a bird’s innate sensory biases to particular song types. This is clearly a much simpler representation of the genetic contribution to song that the note transition matrices used before, but is a simplification that makes the later mathematical analysis of the model tractable. This representation also makes the model more widely applicable to a number of species;

note to note transition biases may be a reasonable model for the munia and the Bengalese finch, and perhaps other species, but as the empirical evidence discussed above shows, the females of other species appear to favour a wide range of song features. Also unlike the model from chapter 3 where I included genes that coded for the SPDs, in this model I do not consider any (direct) genetic contribution to a bird's production biases. Again this makes mathematical analysis considerably simpler, but also seems justified by the result from chapter 3 that the population came to rely only on their auditory copy of song in the SRD as the production system can always be trained using auditory feedback.

This gene affects adult song behaviour in both sexes in the following way. A high quality female with allele A_2 has a bias to acquiring type 2 song, and if she is exposed to this song type as an infant, she will memorise it and later will preferentially mate with a male singing a type 2 song. The strength of this mating preference is represented by the parameter s, with $s = 1$ implying such a female will only mate with males singing type 2 song, and $s = 0$ meaning the female mates indiscriminately, with intermediate values implying intermediate preference for type 2 song. All other females are unable to differentiate type 1 song from type 2 song and so will mate indiscriminately with a male singing either type. A high quality male with allele A_2 which is exposed to type 2 song is assumed to memorise this song and then go on to learn to produce it as an adult, i.e. I am again assuming that the males follow the auditory template model of song development. Other males are only able to develop a type 1 song.

To model song acquisition then, I assume that each infant bird is exposed to a set of m songs, randomly sampled from all the songs of the previous generation. They then pick a single model song to learn from. If the set contains at least one instance of type 2 song a high quality bird with allele A_2 will learn type 2 song. All other birds are unable to learn type 2 song, and so will acquire type 1. According to these assumptions, the probability, P_{T_2}, of any bird hearing at least one instance of type 2 song is given by equation 5.1.

$$P_{T_2} = 1 - (1 - t_2)^m \tag{5.1}$$

Quality (δ)	**Allele** (a_2)	**Model song**	**Adult song**	**Adult preference**	**Label** (X_i)
L	A_1	1	1	either	X_1
H	A_1	1	1	either	X_2
L	A_2	1	1	either	X_3
H	A_2	1	1	either	X_4
L	A_1	2	1	either	X_5
H	A_1	2	1	either	X_6
L	A_2	2	1	either	X_7
H	A_2	2	2	2	X_8

Table 5.1: The effect of the three factors on the adult phenotype

The following three factors thus contribute to an adult bird's phenotype:

- The bird's environmentally specified quality, either *high* (denoted as H) or *low* (denoted as L)
- The allele it inherits from its parents, either A_1 or A_2
- The set of songs the bird hears as an infant

The effects of these factors on the adult song and song preference behaviour is shown in table 5.6. There are 8 possible combinations of quality, allele and model song, giving 8 types of individual of each sex in the model which I label $X_1 \ldots X_8$, as shown in table 5.1. In summary, only high quality individuals bearing allele A_2 and which are exposed to type 2 song, identified as type X_8 in the table, will develop a mating preference for type 2 song (as females) and are able to produce type 2 song (as males). All other individuals show no adult mating preference and can only produce type 1 song. Type 2 song is thus a reliable indicator of male quality; only high quality males are able to acquire and later produce it.

With this model I am interested in the conditions under which allele A_2, which governs the ability to acquire type 2 song, can spread in a population even when it has no effect on some carriers of the allele (namely types X_3, X_4 and X_7), and potentially reduces the number of potential mates for the choosy X_8 females (assuming that $s > 0$). I am also interested in whether cultural transmission of the type 2 song can be stable despite the fact that that it cannot be acquired or produced by every individual in the population.

In the models, I assume that the effect of quality on fitness is to affect the number of viable offspring a mated pair produce. In reality, developmental stress may also affect an individual's chance of survival until reproductive age. However, with this model I am only concerned with the effect of quality on reproductive success, and so assume that birds of high and low quality are equally viable until reproductive age.

This definition of fitness is rather different to that assumed in the models in previous chapters in which fitness was defined as species recognition. Here I assume that both song types can serve equally well to identify conspecifics, instead here I assume that females use the song type to pick between conspecific males according to their quality rather than between conspecifics and heterospecifics.

I define the fitness of mated pairs as a function of the qualities of both the female and her chosen mating partner. The fitness of a bird that does not mate is assumed to be 0. There are four classes of possible matings, a high or low quality female mating with a high or low quality male. The relative fitnesses of each of these classes of pairings are defined in table 5.2. This value can be considered a measure of the relative number of viable offspring that a mated pair will produce, and I assume here that these values will take a value between 0 and 1. Alleles A_1 and A_2 are equally likely to be transmitted from an $A_1 \times A_2$ mating.

	Male	
Female	**H**	**L**
H	1	f_{HL}
L	f_{LH}	f_{LL}

Table 5.2: The relative fitness of each of the 4 possible classes of mated pairs as a function of individual quality

5.5 An agent-based model

The first full model is an agent based model similar to the ones presented in the previous chapters. The model for song is rather simpler though, agents can only sing song type 1 or type 2 which is just represented by a single integer, and the only genetic component is the single allele A_1 or A_2.

Birth The agent's genetic preference is decoded from its genome.

Development The quality of the agent, either high or low, is determined randomly according to the parameter δ. Each agent is exposed to m songs from the previous generation, and picks one model song according to its genetic preference. If the agent carries allele A_2, is high quality and hears at least one instance of type 2 song it will develop an adult preference for type 2 song. If the agent is male it is assumed to use this preference to guide its song learning and will learn to sing a type 2 song as an adult, in a similar fashion to the model of learning developed in chapter 3. All other agents will have no preference and will only be able to develop song type 1.

Adulthood Each agent is exposed to n potential mates and selects a mate based on its adult song preference. X_8 type agents will preferentially pick an agent singing song type 2, i.e. other X_8 agents. The parameter s defines the probability that an X_8 agent will exclusively mate with another X_8 agent, and will not mate at all if an X_8 agent is not found in the n potential mates. Otherwise the X_8 agent, and all other agents, pick indiscriminately from the potential mates. Fitness scores are assigned according to the parameters given in table 5.2. For all experiments described here these were set as follows: $f_{LH} = 0.75$, $f_{HL} = 0.5$, $f_{LL} = 0.25$. The fitness of an agent who does not mate is assumed to be 0.

Reproduction Agents from the population are selected probabilistically according to their fitness score and their genes are subject to a low mutation rate to produce new child agents. This is implemented with a standard genetic algorithm, using tournament selection with a tournament size of 4 for all experiments described here, and a low genetic mutation rate, μ_g. Mutation is modelled by simply flipping the gene from A_1 to A_2 or vice versa. There is no crossover in this model because each agent only has a single atomic allele.

Death Each agent in the population's adult song is sampled and the resulting songs are stored for the next generation to learn from. Each song is subject to a low cultural mutation rate, μ_c, from type 1 to type 2 or vice versa. All of the current birds in the population are removed and their children become the new population.

5.5.1 Experimental results for the agent-based model

5.5.1.1 Experiment 1: A_1 fixation stability with mutation

For the first experiment I created a population of 100 agents all initialised with allele A_1 and set all of the initial songs to type 1. I then let the population evolve for 6000 generations tracking the frequency of A_2, a_2, the proportion of type 2 songs, t_2 and the average fitness of the population. Figure 5.2 shows the average results of running this experiment 100 times.

We see that the A_2 allele quickly invades the population, and t_2 goes up to 0.5. This is the highest possible value for t_2 as the parameter δ is set to 0.5 for this experiment.

5.5.1.2 Experiment 2: A_2 fixation stability with mutation

To test the stability of a population fixated on allele A_2 to invasion from mutations to A_1 I repeated the previous experiment, except that the population was initialised with A_2 at fixation and t_2 was initialised to the value of δ. Figure 5.3 shows that it seems A_2 is stable to invasion from A_1.

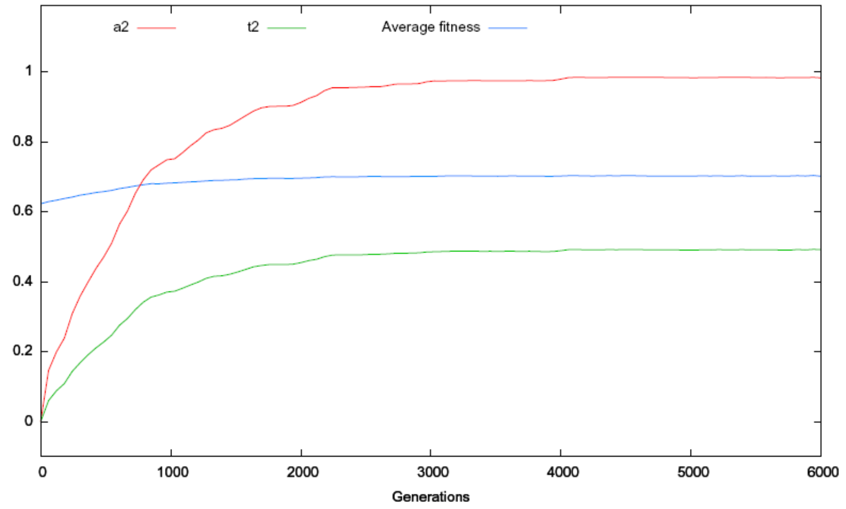


Figure 5.2: Graph showing the frequency of the A_2 allele (a_2), the proportion of type 2 song (t_2) and the average fitness of the population against generations from a population initially fixated on allele A_1. Note that allele A_2 quickly invades the population. Parameters used here: $\delta = 0.5$, $m = 10$, $n = 5$, $s = 0.75$, $\mu_g = 0.01$, $\mu_c = 0.01$

These results suggest that allele A_2 can invade a population fixated in A_2 and that a population fixated on A_2 is stable. The experiments also show that cultural transmission of the less learnable type 2 song can be stable.

This model has a large number of parameters and while we could investigate the effect of each of them with many runs, given the simplicity of the model it is possible to investigate it more generally with a purely mathematical approach, and this is what I present in the next few sections.

5.6 Mathematical models

In the following few sections I attempt to get a deeper understanding of the dynamics of this model of the developmental stress hypothesis with a purely analytical approach. Population genetic models are a classic tool for investigating evolutionary phenomena such as sexual selection and can provide more generality and insight into the model dynamics than an agent based simulation. They do so, though, at the expense of some simplifying assumptions. First, to make the analysis possible, and in keeping with

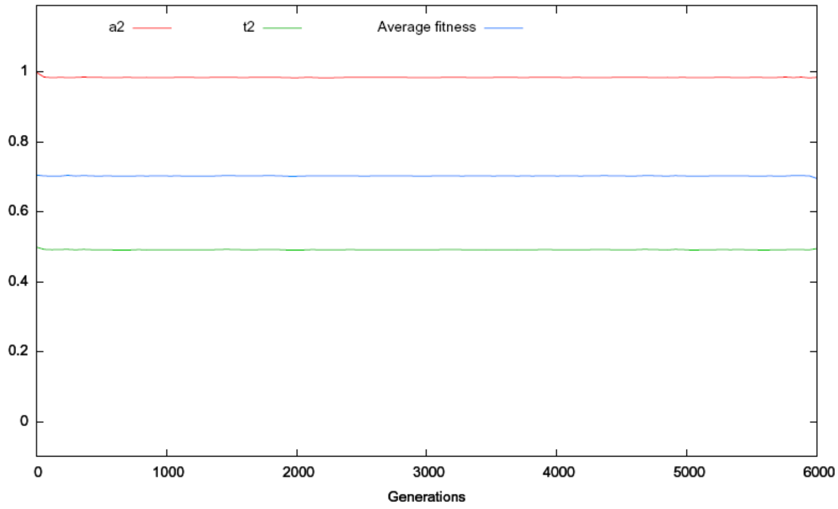


Figure 5.3: Graph showing the frequency of the A_2 allele (a_2), the proportion of type 2 song (t_2) and the average fitness of the population against generations from a population initially fixated on allele A_2. Note that allele A_2 remains at fixation throughout the run. Parameters used here: $\delta = 0.5$, $m = 10$, $n = 5$, $s = 0.75$, $\mu_g = 0.01$, $\mu_c = 0.01$

related models (e.g. Kirkpatrick, 1982; Aoki, 1989; Lachlan and Feldman, 2003), the models assume an infinite population with discrete generations, a 1:1 sex ratio and haploid genetics.

The frequencies of the 8 types are given by the values of δ, a_2 and t_2 in each generation, as shown in table 5.3, which I denote as x_i.

Quality (δ)	**Allele** (a_2)	**Model song** (P_{T_2})	**Adult song**	**Adult preference**	**Label** (X_i)	**Frequency** (x_i)
L	A_1	1	1	either	X_1	$(1-\delta)(1-a_2)(1-P_{T_2}) = x_1$
H	A_1	1	1	either	X_2	$\delta(1-a_2)(1-P_{T_2}) = x_2$
L	A_2	1	1	either	X_3	$(1-\delta)a_2(1-P_{T_2}) = x_3$
H	A_2	1	1	either	X_4	$\delta a_2(1-P_{T_2}) = x_4$
L	A_1	2	1	either	X_5	$(1-\delta)(1-a_2)P_{T_2} = x_5$
H	A_1	2	1	either	X_6	$\delta(1-a_2)P_{T_2} = x_6$
L	A_2	2	1	either	X_7	$(1-\delta)a_2 P_{T_2} = x_7$
H	A_2	2	2	2	X_8	$\delta a_2 P_{T_2} = x_8$

Table 5.3: The frequency of each bird type $X_1 \ldots X_8$ as functions of δ, a_2 and t_2.

5.6.1 Polygyny

	Males			
Females	X_1	$\dots$	X_7	X_8
X_1				$x_1 \cdot x_8$
$\vdots$		$x_i \cdot x_j$		$\vdots$
X_7				$x_7 \cdot x_8$
X_8	$(1-s) \cdot x_8 \cdot x_1$	$\dots$	$(1-s) \cdot x_8 \cdot x_7$	$s \cdot x_8 \cdot P_{X_8}$

Table 5.4: Frequencies of matings for the polygynous model

For the first mathematical model, I assume a polygynous mating system. I am aware that many songbird species are largely monogamous and I model this in section 5.6.2, but as these models can be seen as mathematical variations on the polygynous model, for clarity I first present the polygynous case. This also allows comparison with other models of the evolution of culturally transmitted song which also assume polygyny (e.g. Aoki, 1989; Lachlan and Feldman, 2003).

$$P_{X_8} = 1 - (1 - x_8)^n \tag{5.2}$$

Each time a female chooses a mate she samples n males and will mate with the male whose song she prefers the most. The X_8 females prefer to mate with X_8 males, and the probability they will exclusively pick an X_8 male from their set of n is given by s, if there is not an X_8 male in this set they will not mate. The probability, P_{X_8}, that at least one X_8 male is in the sample of n males is given by equation 5.2.

A proportion $1-s$ of the time, the X_8 females, and all other females pick a mate at random, but will always find a mate. The frequencies of matings between the different types of birds under these assumptions is given in table 5.4. As I am assuming polygyny, each male is able to mate multiple times.

According to these assumptions we can write recursion equations for the value of a_2 in the next generation, $a_2{}'$, and the value of t_2 in the next generation, $t_2{}'$, which is simply the proportion of X_8 males in the current generation. These recursions therefore include both genetic and cultural transmission. In the equations the left hand side of each of the products represents the female types while the right hand side represents the male types. The equations are defined such that for allele A_2 to spread, the average fitness of birds with this allele (w_{A_2}) must be greater than the average fitness of birds with allele A_1 (w_{A_1}).

$$t_2{}' = x_8 \tag{5.3}$$

$$a_2{}' = \frac{w_{A_2}}{w_{A_1} + w_{A_2}} \tag{5.4}$$

where

$$\begin{aligned}
w_{A_2} = f_{LL} \cdot & \Big[(x_3+x_7)(x_1+x_3+x_5+x_7) \Big] + \\
f_{LH} \cdot & \left[(x_3+x_7)(x_4+x_8) + \frac{(x_1+x_5)(x_4+x_8)+(x_3+x_7)(x_2+x_6)}{2} \right] + \\
f_{HL} \cdot & \left[((1-s)\cdot x_8 + x_4)(x_3+x_7) + \frac{(x_2+x_6)(x_3+x_7)+((1-s)\cdot x_8+x_4)(x_1+x_5)}{2} \right] + \\
1 \cdot & \Big[s \cdot x_8 \cdot P_{X_8} + (1-s)\cdot x_8 \cdot (x_4+x_8) + x_4 \cdot (x_2+x_4+x_6+x_8) + \\
& \qquad \frac{(1-s)\cdot x_8 \cdot (x_2+x_6) + (x_2+x_6)\cdot x_8}{2} \Big]
\end{aligned}$$

$$\begin{aligned}
w_{A_1} = f_{LL} \cdot & \Big[(x_1+x_5)(x_1+x_3+x_5+x_7) \Big] + \\
f_{LH} \cdot & \left[(x_1+x_5)(x_2+x_6) + \frac{(x_1+x_5)(x_4+x_8)+(x_3+x_7)(x_2+x_6)}{2} \right] + \\
f_{HL} \cdot & \left[(x_2+x_6)(x_1+x_5) + \frac{(x_2+x_6)(x_3+x_7)+((1-s)\cdot x_8+x_4)\cdot(x_1+x_5)}{2} \right] + \\
1 \cdot & \left[(x_2+x_6)(x_2+x_4+x_6) + \frac{(1-s)\cdot x_8 \cdot (x_2+x_6) + (x_2+x_6)\cdot x_8}{2} \right]
\end{aligned}$$

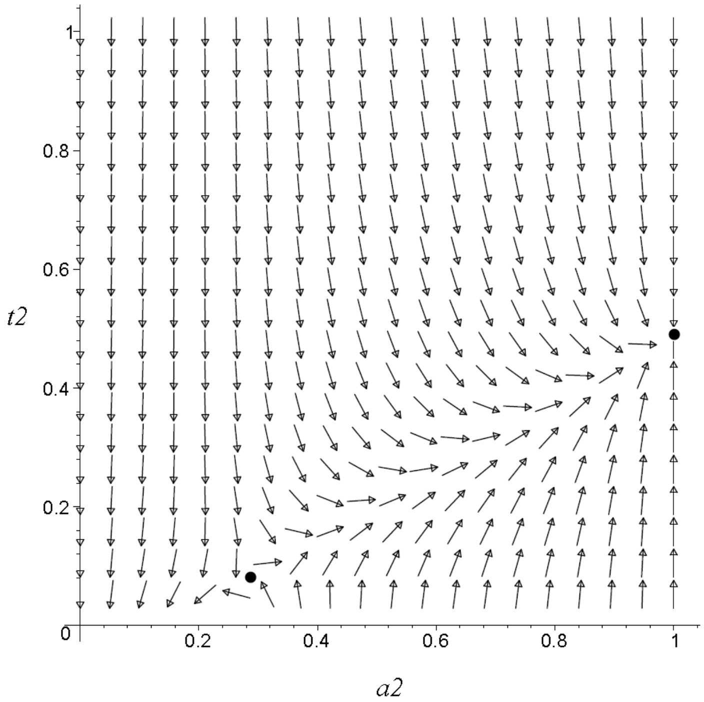


Figure 5.4: Vector field plot showing how the frequency of the type 2 song preference allele, a_2, and the frequency of type 2 song in the cultural environment, t_2, change from a range of possible values in the polygynous model. The equilibrium points are marked with a circle. The parameter values used here are: $\delta = 0.5, m = 10, n = 5, f_{LH} = 0.75, f_{HL} = 0.5, f_{LL} = 0.25$. Note that the maximum possible value of t_2 is given by δ so at the $\hat{a}_2 = 1$ equilibrium in this graph $t_2 = 0.5$

5.6.1.1 Co-evolutionary dynamics

The co-evolution of the A_2 allele and type 2 song can be graphically represented with a vector-field plot in two dimensions which shows how a_2 and t_2 change from a range of possible values, this is shown in figure 5.4. This figure demonstrates that for all values of a_2 when $t_2 \neq 0$ there is frequency-dependent selection for or against allele A_2, except for an internal equilibrium. While the precise value of a_2 above which allele A_2 invades depends on the value of t_2, the vector plots show that it is approximately true that when a_2 is above the value at the internal equilibrium, A_2 invades. When a_2 is below this value, generally A_2 is selected against. Thus I use the value of a_2 at the internal equilibrium as an indicator of the value of a_2 above which A_2 will invade.

5.6.1.2 Model equilibria

Figure 5.4 shows us the expected direction of change of a_2 and t_2 for a wide range of possible values. It is clear from the plot that there are three attractor regions that the system evolves towards and we can investigate these attractor or *equilibrium* points mathematically.

At equilibrium the values of a_2 and t_2 are stable over time, i.e. $a_2{}' = a_2$ and $t_2{}' = t_2$. The recursion equations 5.4 and 5.3 can be solved to give the values at which these conditions hold. The equilibrium equations do not have a general analytic solution (because of terms raised to the powers of the parameters m and n), but we can look at extremes of the model to get some understanding of the general behaviour. If we assume that $t_2 > 0$ and that m is large then the probability of hearing a type 2 song, $P_{T_2} \approx 1$ so we can ignore the terms raised to the power of m. Substituting this into the equations means we find two equilibria where A_2 is either at fixation in the population or is extinct. The equilibrium values for a_2 and t_2 at these points are; $\{\hat{a_2} = 0, \hat{t_2} = 0\}$, $\{\hat{a_2} = 1, \hat{t_2} = \delta\}$. There is also an polymorphic equilibrium where alleles A_1 and A_2 coexist, the values of a_2 and t_2 at this equilibrium are given by equations 5.5 and 5.6 respectively.

$$\hat{a_2} = \frac{e^{\left(\frac{Y}{n}\right)} - 1}{\delta \cdot (e^{\left(\frac{Ym}{n}\right)} - 1)} \tag{5.5}$$

$$\hat{t_2} = 1 - e^{\left(\frac{Y}{n}\right)} \tag{5.6}$$

where

$$Y = ln\left(1 + \frac{f_{HL} \cdot \delta - f_{HL} - \delta}{2}\right)$$

The equilibrium we have found when $\hat{a_2} = 0$ is not actually a valid equilibrium of the full model, as when $a_2 = 0$, $t_2 = 0$, so the assumption that $P_{T_2} \approx 1$ no longer holds. In fact solving the full model with this assumption relaxed finds one solution when $t_2 = 0$, $\{\hat{a_2} = a_2, \hat{t_2} = 0\}$. This means that when $t_2 = 0$ the value of a_2 will not change. This is an intuitive result as when there are no type 2 songs in the environment the birds carrying A_2 behave in exactly the same way as the A_1 carrying birds, and so selection cannot act on the frequency of A_2.

5.6.1.3 Stability analysis

The equilibrium expressions we have found tell us the values at which a_2 and t_2 will not change for particular parameter values. They do not tell us whether the system is likely to evolve towards these values, or how stable these equilibria are to perturbations, e.g. mutation or noise.

We can however also mathematically investigate the (local) stability of the equilibria by finding the leading eigenvalue of the (Jacobian) stability matrix of the model evaluated at each of the equilibrium values (for details of this technique see, e.g., Otto and Day, 2006). This tells us the effect that a small perturbation, ε, will have on the equilibrium values of a_2 and t_2 (ε is assumed to be small enough to ignore in second and higher order terms). If the leading eigenvalue is greater than 1 then the perturbation will grow over time and the system will move away from the equilibrium, and so it is unstable. If the leading eigenvalue is less than 1 the perturbation will shrink over time and the equilibrium is locally stable.

To make algebraic analysis possible I again have to make the simplifying assumption that $t_2 > 0$ and that m is large, so that $P_{T_2} = 1$. In this case for the A_2 fixation equilibrium, where $\{\hat{a}_2 = 1, \hat{t}_2 = \delta\}$, the only non-zero eigenvalue is given by equation 5.7.

$$\lambda = \frac{f_{LL}\cdot(2-4\delta+2\delta^2)+f_{LH}\cdot(2\delta-2\delta^2)+f_{HL}\cdot(2\delta-s\delta-2\delta^2-s\delta^2)+2\delta^2-s\delta^2}{2\cdot[f_{LL}\cdot(1-2\delta+\delta^2)+f_{LH}\cdot(\delta-\delta^2)+f_{HL}\cdot(\delta-\delta^2-s\delta+s\delta^2)+s\delta-s\delta(1-\delta)^n+\delta^2-s\delta^2]} \tag{5.7}$$

Equation 5.7 is a rather complicated expression and I have been unable to find simple expressions of the parameters which allow us to see when $\lambda < 1$, i.e. when the equilibrium will be stable. Instead, to investigate the effect of the various parameters on the eigenvalue I have varied one parameter at a time while holding the others constant and plotted the resulting values of λ, the results of this analysis are shown in figure 5.5.

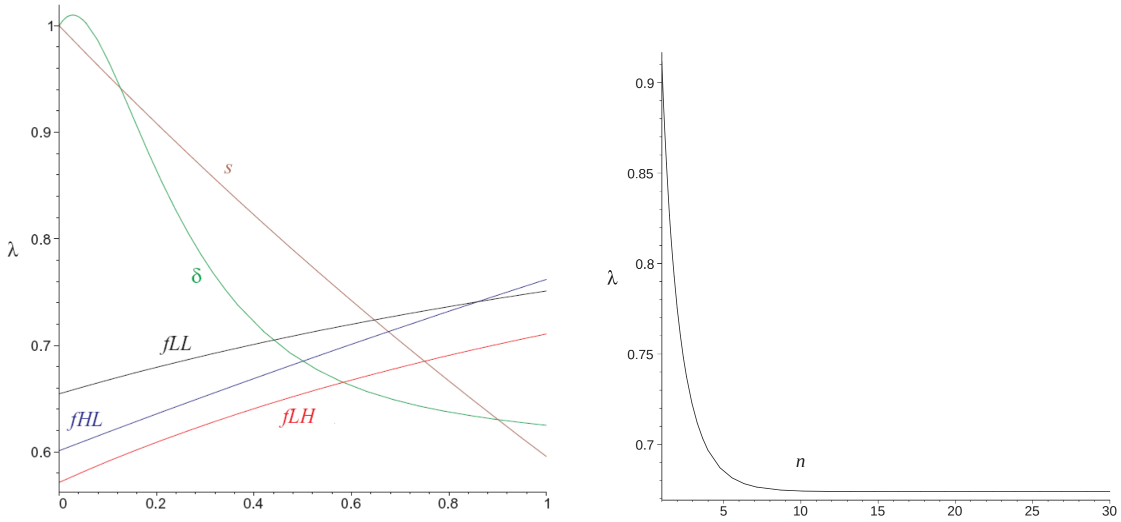


Figure 5.5: Graphs showing the effect of the parameters on the leading eigenvalue of the A_2 fixation equilibrium for the polygynous model. The default parameters values used here are: $\delta = 0.5$, $m = 10$, $n = 5$, $s = 0.75$, $f_{LH} = 0.75$, $f_{HL} = 0.5$, $f_{LL} = 0.25$. Note that $\lambda < 1$ for all parameter values except when $\delta <\sim 0.05$ and $s = 0$.

This plot demonstrates that $\lambda < 1$ for a wide range of parameter values, the only time that $\lambda >= 1$ is when δ is very low, implying that the majority of the population is low quality, or $s = 0$, i.e. the X_8 females are not choosy at all. This suggests that this equilibrium is generally stable for plausible parameter values.

For the polymorphic equilibrium the only non-zero eigenvalue is a very complicated expression, but using a computer algebra system (Maple) I have been able to perform the same graphical analysis as used above. The effects of each of the parameters is shown in figure 5.6. We see that the $\lambda > 1$ for a wide range of parameter values which suggests that this equilibrium is generally unstable.

For the final equilibrium of the model where $\{\hat{a_2} = a_2, \hat{t_2} = 0\}$ and we no longer assume that $P_{T_2} = 1$, there are two non-zero eigenvalues: $\lambda_1 = 1$ and $\lambda_2 = \delta\, a_2\, m$. The first eigenvalue reflects the fact that when $t_2 = 0$ any perturbation in a_2 will persist over

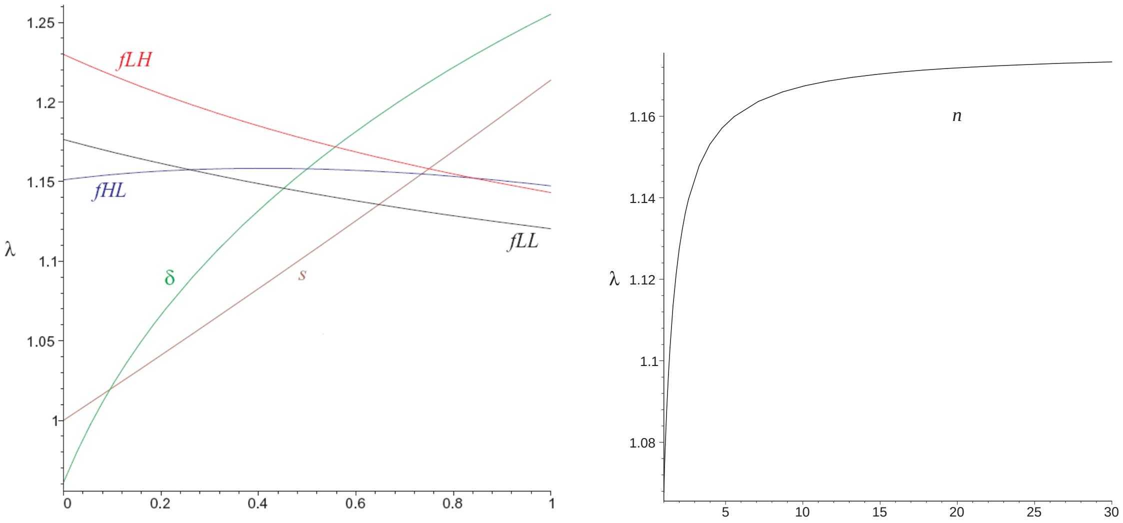


Figure 5.6: Graphs showing the effect of the parameters on the leading eigenvalue of the polymorphic equilibrium for the polygynous model. The default parameters values used here are: $\delta = 0.5$, $m = 10$, $n = 5$, $s = 0.75$, $f_{LH} = 0.75$, $f_{HL} = 0.5$, $f_{LL} = 0.25$. Note that $\lambda > 1$ for a wide range of parameter values except when $\delta <\sim 0.05$.

time but will neither shrink nor grow. The second eigenvalue characterises the stability of a perturbation in t_2. $\lambda_2 > 1$ when $a_2 > \frac{1}{\delta m}$, so if a_2 drifts to a value larger than this a perturbation to t_2 will grow. Therefore, as δ and m increase this threshold value decreases and so the likelihood of invasion by A_2 increases.

5.6.1.4 Consequences for a finite population

The stability analysis shows that the A_2 fixation equilibrium is the only generally stable equilibrium. The polymorphic equilibrium is unstable for a wide range of the parameter values, and the equilibrium where $t_2 = 0$ is stable to perturbations in t_2 for some values of a_2 and the relevant parameters, but perturbations in a_2 will persist. This may have some interesting consequences for a finite population.

When $t_2 = 0$ selection cannot act on a_2. This means that in a finite population the value of a_2 will be controlled by effects such as pleiotropy and drift. If a_2 were to drift to a value greater than the polymorphic equilibrium value and there was then some cultural invention or mutation event which created some type 2 songs, allele A_2 could then invade the population and remain stable[4]. As the stability analysis showed, invasion of allele A_2 is the only globally stable equilibrium value, but the likelihood of invasion in a finite population will increase as the polymorphic equilibrium value of a_2 decreases, i.e. the basin of attraction for the invasion of A_2 gets larger.

The polymorphic equilibrium value of a_2 depends on only 4 of the model parameters m, n, δ and f_{HL}). To demonstrate the effects of these parameters on the polymorphic equilibrium value of a_2, and hence the opportunity for invasion of A_2, I varied one parameter at a time while holding the others constant. The effects of each of these parameters are shown in figure 5.7. These graphs demonstrate that as the number of songs heard in infancy, the number of males each female chooses a mate from (m), and the proportion of high quality individuals increase(δ), the value of a_2 above which A_2 is selected for decreases. As the relative fitness of high quality females who mate with low quality males (f_{HL}) increases it becomes harder for allele A_2 to invade. These results bear out the intuition that as factors increasing the chance that an X_8 female finds an X_8 mate increases, invasion by A_2 is easier. When the relative fitness of high quality females who mate with low quality males increases, the advantage of the X_8 females over other high quality females decreases and so the opportunity for invasion of A_2 decreases.

Some of the model parameters do not affect the equilibrium values, however these parameters do have an effect on the stability of the equilibria, and hence the likelihood of the system evolving to these states. As the graphs in the stability analysis demonstrated, increasing the relative fitness of a low quality female mating with a low quality male (f_{LL}) and that of a low quality female mating with a high quality male (f_{LH}) decreases the stability of invasion by A_2 because larger values reduce the advantage of high quality females. Increasing the strength of the X_8 female's preference for type 2

---

[4] As noted earlier, this is not a completely rigorous analysis as there are some values of a_2 larger than the equilibrium value that will be selected against for low values of t_2, but the value of a_2 at the polymorphic equilibrium gives us a reasonable numerical proxy for values above which A_2 will be positively selected for.

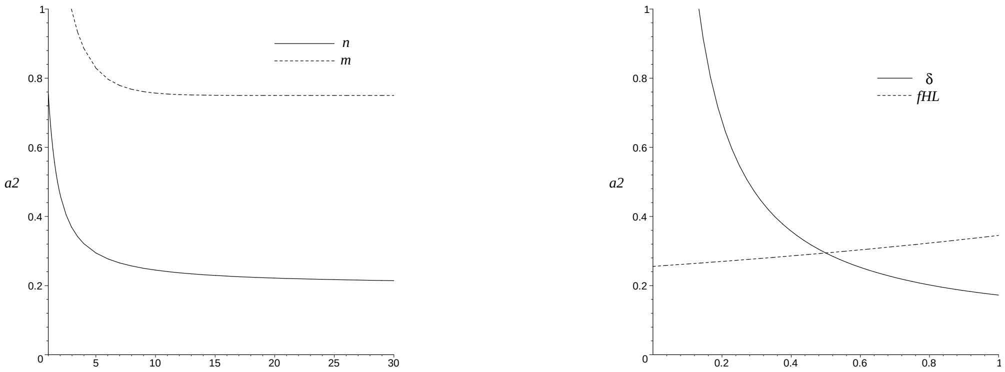


Figure 5.7: Graphs showing the effect of the number of songs heard as an infant (m), the number of males a female chooses a mate from (n), the proportion of high quality birds (δ), and the relative fitness of a high quality female mating with a low quality male (f_{HL}) on the polymorphic equilibrium value of a_2 in the polygynous model. The default parameter values used are: $\delta = 0.5$, $m = 10$, $n = 5$, $f_{HL} = 0.5$. Note that increasing m, n and δ decreases the equilibrium value, making invasion by A_2 more likely, while increasing f_{HL} makes invasion less likely.

song singing males (s) increases the disparity in behaviour between the X_8 females and the other types and so will speed up the invasion or extinction of allele A_2. When $s = 0$ there are no stable equilibria and there is no selection for or against A_2.

5.6.2 Monogamy

Many songbird species are largely monogamous and, as the model assumes that males only provide direct benefits to females (e.g. food, parental care), the assumption in the previous section that high quality males may mate an unlimited number of times may be misleading. It is unlikely that a high quality male would be able to provide such benefits to several females at once. In this section I therefore incorporate monogamy into the model.

To model monogamy we have to assume that as a female chooses a mate this male is no longer available for mating with any other female. When $s > 0$ the X_8 females preferentially choose X_8 males and when they make their choices they will skew the distribution of available X_8 males in the population, all other females choose at random and so they will not change the distribution of males. I have been unable to deal with this mathematically in general and instead have made two simplifying assumptions to allow me to calculate the mating frequencies. Firstly I follow O'Donald (1980) in that I only consider the two cases when either the choosy $s \cdot x_8$ proportion of the X_8 females get to pick their mates *before* all the other females, or when these females choose *after* all the other females[5]. The former approach is the one taken by some previous models of monogamy (e.g. Andersson, 1986), but I model both extremes as these represent the highest and lowest possible number of matings involving X_8 females. These extremes therefore give us an upper and lower bound on the case where the X_8 females choices are intermixed with the other females. The second simplification I make is that the number of mates females choose from, $n = \infty$ so that a choosy X_8 female is guaranteed to find an X_8 male if there are any left in the population (as again otherwise I would have to deal with a changing distribution of male types). I realise that both of these assumptions are biologically implausible, but hope that these extremes of the model will provide some insight on the full behaviour that could not otherwise be obtained.

The frequencies of matings between all 8 phenogenotypes when the X_8 females choose first is given in table 5.5, and the frequencies when the X_8 females choose last is given in table 5.6.

	Males			
Females	X_1	$\dots$	X_7	X_8
X_1				$\frac{x_1 \cdot (1-s) \cdot x_8}{1-(s \cdot x_8)}$
$\vdots$		$\frac{x_i \cdot x_j}{1-(s \cdot x_8)}$		$\vdots$
X_7				$\frac{x_7 \cdot (1-s) \cdot x_8}{1-(s \cdot x_8)}$
X_8	$\frac{(1-s) \cdot x_8 \cdot x_1}{1-(s \cdot x_8)}$	$\dots$	$\frac{(1-s) \cdot x_8 \cdot x_7}{1-(s \cdot x_8)}$	$\frac{((1-s) \cdot x_8)^2}{1-(s \cdot x_8)} + (s \cdot x_8)$

Table 5.5: Frequencies of matings for the monogamous model when the choosy $s \cdot x_8$ proportion of the X_8 females choose before the other females

	Males			
Females	X_1	$\dots$	X_7	X_8
X_1				$x_1 \cdot x_8$
$\vdots$		$x_i \cdot x_j$		$\vdots$
X_7				$x_7 \cdot x_8$
X_8	$(1-s) \cdot x_8 \cdot x_1$	$\dots$	$(1-s) \cdot x_8 \cdot x_7$	$((1-s) \cdot x_8 \cdot x_8) + (s \cdot x_8 \cdot x_8)$

Table 5.6: Frequencies of matings for the monogamous model when the choosy $s \cdot x_8$ proportion of the X_8 females choose after the other females

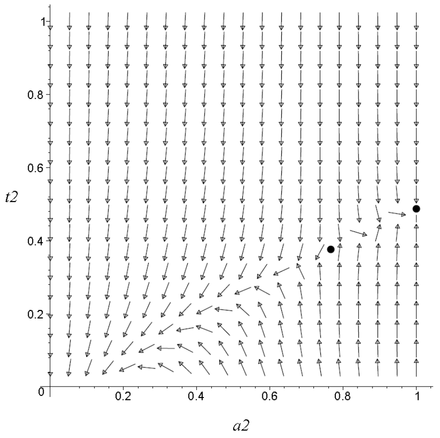


Figure 5.8: Vector field plot for the monogamous model when the X_8 females choose after the other females. The equilibrium points are marked with a circle. The parameter values used here are: $\delta = 0.5$, $m = 10$, $f_{LH} = 0.75$, $f_{HL} = 0.5$, $f_{LL} = 0.25$

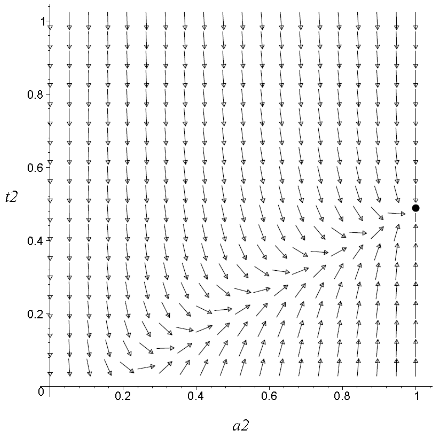


Figure 5.9: Vector field plot for the monogamous model when the X_8 females choose before the other females. The equilibrium point is marked with a circle. The parameter values used here are: $\delta = 0.5$, $m = 10$, $f_{LH} = 0.75$, $f_{HL} = 0.5$, $f_{LL} = 0.25$

5.6.3 Monogamous model 1: The X_8 females choose last

For the case where the X_8 females choose after the other females, a proportion $(1 - x_8) + (1-s) \cdot x_8$ of the X_8 males will have paired off with other females. This will leave a proportion $s \cdot x_8$ of the X_8 males, so the frequency of choosy $X_8 \times X_8$ matings will be $s \cdot (x_8 \cdot x_8)$ and the remaining proportion $s \cdot (1 - x_8)$ of the X_8 females will not find a

---

[5] I do not consider the proportion $(1-s) \cdot x_8$ of the X_8 females as they choose indiscriminately and therefore do not affect the distribution of males.

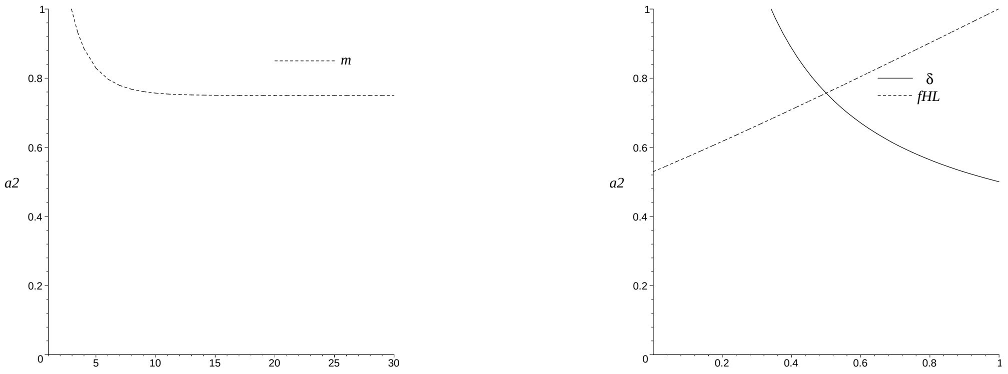


Figure 5.10: Graphs showing the effect of the number of songs heard as an infant (m), the proportion of high quality birds (δ), and the relative fitness of a high quality female mating with a low quality male (f_{HL}) on the polymorphic equilibrium value of a_2 in the monogamous model when the X_8 females choose after the other females. The default parameter values used are: $\delta = 0.5$, $m = 10$, $f_{HL} = 0.5$. Note that increasing m and δ decreases the equilibrium value, making invasion by A_2 more likely, while increasing f_{HL} makes invasion less likely.

mate. This means that this model is in fact a special case of the polygynous model for $P_{X_8} = x_8$ (which is mathematically equivalent to setting $n = 1$) and so the recursion equations 5.4 and 5.3 and the equilibria found earlier are the same for this model, but with n set to 1. A vector-field plot for this model is shown in figure 5.8.

As this model is a special case of the polygynous model, the models shares the same equilibrium values and so the same stability analysis applies here.

We see that the internal equilibrium for this model (with the same values of m, δ and f_{HL} as used in the polygynous model) requires a higher value of a_2 and so we would expect that allele A_2 would be less likely to invade this population than for the polygynous model. Figure 5.10 shows the effect of m, δ and f_{HL} on the polymorphic equilibrium value of a_2, again demonstrating that higher values of m and δ increase the chance of invasion of A_2. Increasing f_{HL} again decreases the chance of invasion by A_2 but it has a rather stronger effect here than in the polygynous model. We can see that when $f_{HL} = 1$ there is no region of selection for A_2 and so it would be very unlikely to invade. In other words, as expected, there needs to be some advantage for high quality females to mate with high quality males.

5.6.4 Monogamous model 2: The X_8 females choose first

For the case where the choosy X_8 females choose before the other females, note that, by definition, the number of X_8 males equals the number of X_8 females, so there will always be enough X_8 males for the choosy X_8 females. As I assume that $n = \infty$, this means that the proportion $s \cdot x_8$ of the choosy X_8 females are guaranteed to be paired with an X_8 male so there remains only a proportion $1-(s \cdot x_8)$ of the X_8 males remaining for the other females to choose. I therefore need to redefine the recursion equations slightly for this model. Essentially I reduce the frequency of any mating event that includes an X_8 male mating with any female type other than X_8 by $1-s$, and I then divide each of the remaining mating event frequencies by $1-(s \cdot x_8)$. The full recursions for this model are given by equations 5.8 and 5.9. A vector-field plot for this model is shown in figure 5.9.

5.6.4.1 Model equilibria

Solving the recursion equations for this model for the equilibrium conditions in the same manner as before finds two equilibria; an A_2 fixation equilibrium where $\hat{a_2} = 1, \hat{t_2} = \delta$, and an equilibrium with no type 2 song where $\hat{a_2} = a_2, \hat{t_2} = 0$. Unlike the two previous models there is no region of selection against A_2 and there is no polymorphic equilibrium.

Although there is no region of selection against A_2 in this model there is a region where there is no selection for A_2. As the stability analysis below demonstrates, the value of a_2 that is required for positive selection for A_2 is given by $a_2 > \frac{1}{\delta m}$. When this condition holds, a small perturbation to t_2 will allow A_2 to be positively selected for. Overall in this model then, we would expect allele A_2 to be able to invade this population more easily than in the previous two models, but the likelihood of invasion is dependent on the values of δ and m, with larger values making invasion more likely.

$$t_2{}' = x_8 \tag{5.8}$$

$$a_2{}' = \frac{w_{A_2}}{w_{A_1} + w_{A_2}} \tag{5.9}$$

where

$$\begin{aligned}
w_{A_2} = f_{LL} \cdot & \left[\frac{(x_3+x_7)(x_1+x_3+x_5+x_7)}{1-(s \cdot x_8)} \right] + \\
f_{LH} \cdot & \left[\frac{(x_3+x_7)(x_4+(1-s)\cdot x_8)}{1-(s\cdot x_8)} + \frac{(x_1+x_5)(x_4+(1-s)\cdot x_8)+(x_3+x_7)(x_2+x_6)}{2\cdot(1-(s\cdot x_8))} \right] + \\
f_{HL} \cdot & \left[\frac{((1-s)\cdot x_8+x_4)(x_3+x_7)}{1-(s\cdot x_8)} + \frac{(x_2+x_6)(x_3+x_7)+((1-s)\cdot x_8+x_4)(x_1+x_5)}{2\cdot(1-(s\cdot x_8))} \right] + \\
1 \cdot & \left[s\cdot x_8 + \frac{(1-s)\cdot x_8\cdot(x_4+(1-s)\cdot x_8)+x_4\cdot(x_2+x_4+x_6+(1-s)\cdot x_8)}{1-(s\cdot x_8)} + \right. \\
& \left. \frac{(1-s)\cdot x_8\cdot(x_2+x_6)+(x_2+x_6)\cdot(1-s)\cdot x_8}{2\cdot(1-(s\cdot x_8))} \right]
\end{aligned}$$

$$\begin{aligned}
w_{A_1} = f_{LL} \cdot & \left[\frac{(x_1+x_5)(x_1+x_3+x_5+x_7)}{1-(s \cdot x_8)} \right] + \\
f_{LH} \cdot & \left[\frac{(x_1+x_5)(x_2+x_6)}{1-(s\cdot x_8)} + \frac{(x_1+x_5)(x_4+(1-s)\cdot x_8)+(x_3+x_7)(x_2+x_6)}{2\cdot(1-(s\cdot x_8))} \right] + \\
f_{HL} \cdot & \left[\frac{(x_2+x_6)(x_1+x_5)}{1-(s\cdot x_8)} + \frac{(x_2+x_6)(x_3+x_7)+((1-s)\cdot x_8+x_4)\cdot(x_1+x_5)}{2\cdot(1-(s\cdot x_8))} \right] + \\
1 \cdot & \left[\frac{(x_2+x_6)(x_2+x_4+x_6)}{1-(s\cdot x_8)} + \frac{(1-s)\cdot x_8\cdot(x_2+x_6)+(x_2+x_6)\cdot(1-s)\cdot x_8}{2\cdot(1-(s\cdot x_8))} \right]
\end{aligned}$$

5.6.4.2 Stability analysis

The only non-zero eigenvalue of the A_2 fixation equilibrium is given by equation 5.10. This is again a rather complicated expression and so I provide a graphical analysis of the effect of different parameter values in figure 5.11. We see again that this equilibrium appears to be stable over a broad range of parameter values.

The eigenvalues for the $\{\hat{a_2} = a_2, \hat{t_2} = 0\}$ equilibrium of this model are the same as for the polygynous case and so the same analysis applies here.

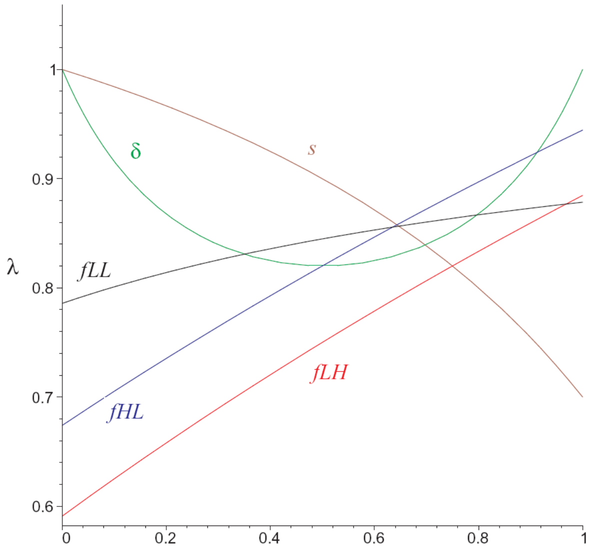


Figure 5.11: Graph showing the effect of the parameters on the leading eigenvalue of the A_2 fixation equilibrium for the monogamous model when the X_8 females choose before the other females. The default parameters values used here are: $\delta = 0.5$, $m = 10$, $f_{LH} = 0.75$, $f_{HL} = 0.5$, $f_{LL} = 0.25, s = 0.75$. Note that $\lambda < 1$ for all parameter values tested here (except when $\delta = 0$ or 1 and when $s = 0$).

$$\lambda = \frac{f_{LL}\cdot(2-4\delta+2\delta^2)+f_{LH}\cdot(2\delta-s\delta-2\delta^2+s\delta^2)+f_{HL}\cdot(2\delta-s\delta-2\delta^2-s\delta^2)+2\delta^2-2s\delta^2}{2\cdot[f_{LL}\cdot(1-2\delta+\delta^2)+f_{LH}\cdot(\delta-s\delta-\delta^2+s\delta^2)+f_{HL}\cdot(\delta-s\delta-\delta^2+s\delta^2)+s\delta+\delta^2-2s\delta^2]} \tag{5.10}$$

For both of these monogamous models, the parameters f_{LL}, f_{LH} and s have the same effect on the equilibria as we found for the polygynous model.

5.7 Discussion

The developmental stress hypothesis proposes that learned features of male song can be used as an indicator of early developmental stress. The models presented here demonstrate 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. They also show that there can be stable cultural transmission of these less learnable song types. The opportunity for invasion of such an allele in a finite population depends, however, on several biologically relevant parameters, including the prevalence of stressors, the number of unique songs heard as an infant and the number of potential mates available to females to choose from. Assumptions about the mating system also have an important effect. The models also incorporate some considerations that have yet to be investigated empirically, such as the the effect of early song exposure and developmental stress on female choice.

In the first two mathematical models we found a strong threshold effect, due to the frequency-dependent selection for or against X_8 individuals. We found that the frequency of the A_2 allele must rise above a threshold value, by some mechanism such as random drift, before it can stably invade the population. How likely this situation is to arise is therefore significant in analysis of the model. All I require of such an allele is that it makes the bearer predisposed to learning some song types rather than others; this might be as a result of some form of auditory bias that occurred by mutation or drift, or by selection on auditory capacities not related to song preference. I feel that such mutations are likely to have occurred fairly frequently in the evolution of the songbirds to give rise to the great variety of song types we find today. The only special property I require of this bias is that the ability to learn a song that satisfies the bias is correlated in some way with the level of developmental stress. As discussed in section 5.6.1, 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. High quality bearers of the allele are then able to reliably find a high quality mate and, in sufficient numbers, the allele can stably invade the population.

I make very few assumptions about the nature of the song types, so any feature of song that has the properties I assume here may be used as an indicator; the key property is that an individual's ability to learn the feature is affected by developmental stress. This seems consistent with the biological data as it appears that females of several different species attend to different features of song, but importantly many of which appear to be affected by developmental stress. The DSH may thus provide a unifying explanation for the wide variety of female preferences observed in the songbirds.

It is a surprising result of the models that selection can favour a preference for a song type that is less learnable, and that such types can be stably culturally transmitted. This seems relevant to the discussion in the previous chapter, section 4.3.2, on song complexity. Several types of song that are preferable to females have been described by researchers as ‘complex’, but this term is frequently defined differently for different species; from syllable repertoire size in the great reed warbler to song linearity in the Bengalese finch. The DSH suggests a unifying, empirically grounded, definition of song complexity as a feature of song that is only learnable by a high quality individual. Such song ‘complexity’ is selected for because it allows females to differentiate among males according to their learning ability, while ‘simpler’ forms do not. The idea that songs that are less learnable might be more attractive to females is supported by recent work showing that both female swamp sparrows (Ballentine et al., 2004) and canaries (*Serinus canaria*) (Draganoiu et al., 2002) prefer songs that are closer to the physical performance limit of males (in these cases of the trill rate and frequency bandwidth), and therefore hardest to produce and, plausibly, to acquire. Vallet and Kreutzer (1995) show that different song phrases produced by male canaries produce distinct responses in females. The reason the females prefer these so-called ‘sexy syllables’ has not been definitively established, but Vallet et al. (1998) suggest that these syllables may provide information about male condition. It would be interesting to investigate if these syllables are harder for young canaries to acquire and therefore may be another possible indicator of learning ability.

While I have presented the model as a model of the evolution of particular genetic learning biases and culturally transmitted song types, under a more general interpretation the model may also be applied to the origin of song learning. Individuals with allele A_1 have a fixed preference for type 1 song and, as males, will go on to develop type 1 song regardless of their quality or the songs they hear as infants. Under this new interpretation these individuals may be viewed as having genetically specified song. Developmental stress has no effect on individuals bearing this allele and so females cannot use song to identify high quality males. The alternative allele A_2 may now be considered as coding for the ability to acquire song culturally, along with a preference for type 2 song which can only be transmitted culturally. If they are only exposed to type 1 song they will acquire this and will function identically to bearers of allele A_1, however if there is type 2 song in the cultural environment high quality individuals will develop an adult preference for type 2 song and the males will learn to produce this song type. Low quality individuals with this allele are assumed to be unable to acquire the more complex type 2 song and so ‘fall back’ to type 1 song. Thus only bearers of allele A_2, the ‘learning’ allele, are affected by developmental stress, and only females who are capable of cultural transmission are able to differentiate males according to their song. The DSH may therefore also be able to explain the origin of song learning, as well as identifying its role as an indicator mechanism. However, while a difference in song preference may reasonably (I submit) be modelled by a mutation at a single locus, it seems rather implausible that the entire song learning system could result from a single mutation, and to demonstrate this argument convincingly would require substantial changes to the model. The same problem holds, however, for Aoki’s (1989) model, and for such a wider interpretation of Lachlan and Feldman’s (2003) models.

5.7.1 Predictions

The models presented here are admittedly simple, and there are a number of modifications that could be made to make them more biologically plausible, such as working with a finite population and modelling the cultural transmission of song in a more realistic way than the panmictic cultural environment I assume here. Making such modifications would, however, make the mathematical analysis of the interactions of the parameters considerably harder and less clear. Nevertheless, I feel that these simple models incorporate some of the key features of the DSH and as they stand, the models suggest a number of predictions which may be testable empirically and I present some here.

Firstly, if there are a low proportion of high quality birds, i.e. the average level of developmental stress in the population is high, the model predicts that we will be less likely to see such a system evolve. This is because it will be harder for the few high quality birds to overcome the frequency-dependent selection against them below the threshold value (which I discuss further below). The model also predicts that we are more likely to see such a system evolve in species where the females are able to pick mates from a relatively large pool of males, and, in monogamous species, where the high quality females are able to pick mates earlier in the breeding season than low quality females (as originally suggested by Darwin, 1879). A final, rather intuitive, prediction is that if the fitness detriment of mating with a low quality male is small then there is little advantage to finding a high quality male and such a system is much less likely to evolve.

5.8 Wider implications

We have seen that a number of empirical studies support the developmental stress hypothesis and the models developed in this chapter lend further support. While both the original hypothesis and the models presented here were designed to apply to song learning, there is little built in that constrains them to songbirds. Wherever we can show that similar initial assumptions hold the results may apply.

5.8.1 Developmental stress and human language acquisition

The DSH attempts to provide an adaptive explanation of song learning and this idea stemmed from two observations of songbird life history; that songbirds are particularly susceptible to developmental stress which may have long-term fitness relevant effects, and that song learning occurs during this potentially stressful period and may thus act as a window on early development. In trying to provide an explanation of the adaptive functions of vocal learning in humans we may be able to apply a similar argument and I provide some initial evidence supporting such an argument here.

Human newborns are generally accepted to be somewhat altricial; at birth they are essentially immobile and require extensive support from adults to protect and nourish them for several years. Brain growth and development is very significant during this period (e.g. Locke and Bogin, 2005). Galler and Barrett (2001) demonstrate that malnutrition experienced by young children may have lifelong consequences that are not reversed by later adequate nutrition. They performed a longitudinal study in Barbados comparing individuals aged 28-34 years who experienced malnutrition in their first year of life with control individuals from the same class at school. They found that early malnutrition had a long-term effect on behavioural and cognitive development and on later IQ. This suggests that early developmental stresses may have long-term fitness relevant effects on adult humans.

Human children also begin to acquire language during the first few years of life and it seems plausible that stresses suffered during this period may affect their language development. Indeed, Cravioto and DeLicardie (1972) show that children that have recovered from severe malnutrition lag behind controls in language development and argue that poor micro-environmental conditions are not sufficient to fully explain the behavioural lag. This means that malnutrition may have a long term and specific impact on language development.

These two studies suggest that the two observations that initially informed the DSH may also apply to humans. Vocal learning in our own species may also therefore be used as an indicator mechanism of early development, which may be useful in mate choice and potentially other contexts. There are though a range of remaining empirical questions that need to be addressed to test this hypothesis fully. For example, we would have to demonstrate that in a context such as mate choice individuals rate a potential partner according to language abilities that have been shown to be affected by developmental stress.

I am not suggesting that providing an indicator mechanism for developmental stress is the only current adaptive function of vocal learning in humans but simply that it may have played such a role at some point in the evolution of our language faculty.

5.8.2 Beyond vocal learning: generalising the DSH

I believe that the core contribution of the DSH is that it proposes a new role for cultural transmission in sexual selection; that the acquisition and possession of particular cultural traits may serve as a reliable indicator of phenotypic quality. This observation may also apply to other culturally transmitted traits that play a role in contexts such as mate choice. While in bird song it appears that generally it is the males who display and the females who make mating decisions based on this display, this need not be the case for us. In humans both males and females play an active role in mate choice and so both sexes may use adult language abilities to decide on a sexual partner.

Human musical ability is another domain where these sorts of arguments may apply. Scholars from Darwin onwards, (e.g. Darwin, 1879; Miller, 2000), have proposed that human musical behaviour may be used as a display in sexual selection context, and Miller (2000) suggests that musical ability may serve as some form of fitness indicator. These studies generally do not, however, consider the key role of cultural transmission in such abilities. To become a proficient musician in a given musical cultural tradition, such as western classical music or in American jazz, requires a significant investment of time and economic resources, and so proficiency in such a tradition provides hard-to-fake evidence that an individual has, or at least had, such resources available.

In providing an explicit mechanistic link between phenotypic condition and learning ability, the DSH may provide a more satisfying justification of these sorts of hypotheses in a range of behaviours affected by cultural transmission.

5.9 Conclusions

The models presented in this chapter synthesise the available biological data and verbal arguments for the developmental stress hypothesis and provide the first theoretical support for a novel role for cultural transmission in sexual selection. As I have argued, this may have implications for other species that rely on cultural transmission, including humans. It is possible that an effect similar to the one investigated here may have played a role at some stage in the evolution of the human language faculty.

The results presented here supplement the results from the previous chapters in that they suggest a plausible and widely applicable adaptive function for song learning. They also provide an satisfying account of why females seem to prefer features of song that we perceive as complex in some way. Indeed, Soma et al. (2006) show that song linearity in Bengalese finches is correlated with rearing conditions in such a way that suggests that birds who sing less linear songs were raised in superior conditions. This result grounds Okanoya's definition of complexity (discussed in the previous chapter) and gives more support to his hypothesis that female choice played an important role in song complexity in the Bengalese finch by linking female preference and male quality.

Chapter 6

Conclusions

In this thesis I have presented a series of formal models that have investigated several issues in the evolution of complex, culturally transmitted communication systems. In this concluding chapter I identify the main contributions of the thesis to an understanding of the origin of vocal learning and vocal complexity in humans and possibly other species. The evolution of these abilities was a significant transition in the evolution of language, but the picture is still far from complete and I identify some questions which still remain to be answered and which should be addressed with future work.

6.1 Main contributions

In this section I identify the main contributions of this thesis to the investigation of the evolution of the two features of human speech discussed in chapter 2; cultural transmission, in the form of vocal learning, and vocal complexity. I also set the results of the work presented here in the wider context of work on the evolution of language.

6.1.1 Vocal learning

Chapter 3 presented a simple computational implementation of the auditory template model, the most widely accepted model of song learning. This implementation provides a concrete instantiation of each of the various components proposed by the ATM which aids in understanding how the various components may interact. This may be useful in future studies investigating further issues in the evolution of this important component of the song system. The proposed implementation is also shown to align with recent empirical studies of the neural implementation of song learning in songbirds.

The chapter then embeds this implementation of song learning in an evolving population of agents where the innate song perception and production biases, and the song system plasticity values, are under genetic control. The results of experiments performed with this model demonstrate that a vocal learning system is only likely to evolve if cultural transmission is reliable enough, and only in species where song is not required until adulthood. These results should hold generally over any proposed function of song learning.

These results suggest that vocal learning in humans is likely to have evolved in an environment where complete language abilities were not required early in life. It is possible that, as with bird song, human vocalisation originally served in a sexual context and such vocalisation would not therefore be required until sexual maturity. It is clear though that although language acquisition still takes a significant period of time, children today are able to speak well before puberty. As I discuss below, although human speech may once have served a similar function to song, it is now used in a number of non-sexual contexts, and in particular for the transmission of rich semantic information. Such new functions of speech may be useful early in life and so vocal learning abilities in humans have adapted to fit these new requirements.

The model presented in chapter 4 demonstrates that a change in the fitness landscape of a species that results in some form of masking of selection, such as domestication, may result in increased influence from the cultural environment. This effect was found to occur because genes that were previously held in check by natural selection were allowed to drift and previously minor influences, such as early acoustic experience, came to have more of an effect on the adult phenotype. It is possible that some change in the ecology of early humans may have had a similar masking effect on our communication system which resulted in the diversification of our vocal abilities. Evidence for such a process may be found in comparative studies of the differences in vocal communication in humans and non-human primates, in particular we should look for signs of genetic drift in the genes underlying our vocal abilities.

The models of the developmental stress hypothesis developed in chapter 5 provide the first theoretical support for a novel evolutionary function for vocal learning. The models demonstrate how vocal learning may serve as an indicator mechanism, permitting females to select a mate using his adult vocalisations as a window on early development. A surprising conclusion of these models when compared to other models of the evolution of cultural transmission is that selection can favour a culturally transmitted song type that is, by definition, less learnable. I also discuss some initial empirical evidence that suggests that developmental stress could have an effect on adult language abilities. Whether this information is used in human mate choice is an outstanding empirical, and testable, question. This argument could be supported further with empirical work on the effect of developmental stress on adult language abilities. Empirical studies could also establish if competence in some features of language that can be shown to be hard to acquire have any measurable effect on human mate choice.

6.1.2 Complex vocalisation

The model presented in chapter 4 shows that an apparent increase in the ‘complexity’ of a vocal communication system may occur without selection for such complexity, but instead as a result of the weakening of genetic biases and the subsequent amplification of latent variation in cultural transmission. This result also highlights the difficulty of classifying observed behaviour as ‘complex’ in a quantifiable and meaningful way. It may be that ‘simple’ behaviours are harder to maintain in the face of mutation pressure, and that apparent complexity does not necessarily require an adaptive explanation.

The approach to modelling song learning as the induction of a finite state machine may find applications beyond modelling work. Until we have a better understanding of the neural implementation of song learning in species with apparently complex syntax such as the Bengalese finch, selection of grammars by minimum description length seems like a principled and reasonable model of learning. The measures used in Teal and Taylor's (2000) approach, such as the grammar encoding length (GEL) may provide a principled way to quantify song complexity in species where finite state syntax seems like a reasonable model of song structure.

A key assumption in the model of the developmental stress hypothesis is that some song types, or features of song, are in some sense harder to learn for lower quality individuals. The ability to produce such song types can then reliably identify high quality individuals. This suggests an empirically grounded and testable definition of song complexity as songs that are harder to learn.

The results of both of these models suggest that complex vocalisation in humans could have evolved without a selection pressure for expressivity, either due to the diversification of behaviour that could result from the masking of selection pressure, or because vocal complexity may serve as a fitness indicator. Once an ability for complex vocalisation, whether it be the result of either of these processes, evolved this increased the information carrying capacity of the vocal signal. Such information carrying capacity could then be co-opted for the transmission of complex semantic information, as we observe in human speech today.

6.1.3 Wider implications

As well as addressing the evolution of these two specific features of human speech, I believe that the work presented in this thesis also addresses some wider issues in the study of the evolution of language which I briefly discuss here.

Firstly the models emphasise the importance of considering the interactions between the three dynamic processes that Kirby (2007) argues contribute to language evolution; namely genetic evolution, individual development and cultural evolution. All of the models I present include some formalisation of each of these processes and the results rely heavily on the interplay between them: for example the model of the developmental stress hypothesis investigates the evolution of a genetic bias to learning particular culturally transmitted traits. The conclusions we obtained from this model could not have been reached without considering the combined effect of each of these three processes.

There has been a long debate about which sorts of processes can explain the nature of human language, these range from Chomsky's (1988) argument that language may not be explained by natural selection but by fundamental physical principles, to Pinker and Bloom's (1990) arguments that language is the result of 'boring' biological evolution, to Kirby's (1999) arguments that many features of language may be explained by purely cultural processes. In the models I have presented the processes of evolution, development and cultural transmission are inextricably linked and I suggest that they must all be taken into account simultaneously for a complete account of language evolution. Studies of the 'innate' component of song structure in birds show us that there is great variation between species in what genetic biases songbirds bring to song acquisition and how these affect adult song. We are unlikely to be able to perform such detailed studies on humans due to obvious ethical issues and we must therefore be very cautious in what we attribute to genes *or* culture in human linguistic behaviour without careful empirical work.

The models also all investigate different ecological conditions that may have an effect on the evolution of a complex, culturally transmitted communication system in birds; including life history, domestication and mate choice. The models show that each of these considerations may play important and sometimes surprising roles in shaping the communication system. Human language is used in many more contexts than bird song and this suggests that even more evolutionary forces may have affected its evolution. Again I suggest that looking for a single adaptive function of language, be it proposition transmission (e.g. Pinker and Bloom, 1990), sexual selection (e.g. Miller, 2000), social 'grooming' (e.g. Dunbar, 1996) or any of the other many proposed functions for language is unlikely to be fruitful. We must take a wider view and accept that language has been, and continues to be, subject to many and varied selective forces. Experiments investigating the role of some proposed function for language can and should continue to be performed, but researchers must be aware that as humans developed the functions of language may have changed and that other functions may also be simultaneously relevant.

Overall then, the work presented here suggests that in trying to understand the evolution of language we must bear in mind the range of processes that gave rise to it and the varied selection pressures under which it evolved. This makes study of language evolution a significant intellectual challenge, but I believe that simple formal models such as those presented here, combined with comparative work on similar communication systems in other species provides us with the most promising avenues for further study. As I argued in chapter 2, bird song is a particularly useful model for language because of the many similarities in its development and, to some extent, implementation, and because it provides us with a convenient model system for culturally transmitted communication. The amount and range of experimental data on the physiology, development and functions of bird song provide a valuable resource for research in language evolution.

6.2 Outstanding issues

As discussed in the introduction, the fact that human language is transmitted culturally as well as biologically is not a simple difference in the mechanism of transmission, it introduces a unique dynamic to the evolution of this complex behaviour. As I demonstrated in chapter 2, this is also true of a small number of other animal communication systems, including bird song. However, it appears that humans are able to use their complex, culturally transmitted communication system for considerably richer communication than birds and other vocal learning species. In particular we use complex signals to convey arbitrarily complex information to each other. Even if, as I have argued, culturally transmitted complex vocal signalling originally evolved in humans for similar functions as in songbirds, at some point we acquired the ability to attribute meanings to these signals. This is therefore another important transition that must be addressed in an account of the evolution of language.

There is little evidence that birds attribute complex meaning to song, but there is evidence that our closest extant relatives the chimpanzee may attribute some form of semantics to vocalisations. Slocombe and Zuberbuhler (2005) show that chimpanzees produce acoustically distinct ‘rough grunts’ when encountering different types of food. One individual apparently understood the differences in vocalisations and was able to use the information encoded in them to guide his search for food. An ability to attribute meanings to vocal signals may therefore have been present in the last common ancestor of humans and chimpanzees.

In two recent related papers (Merker and Okanoya, 2007; Okanoya and Merker, 2007) propose a possible way in which complex vocalisations could come to encode semantic information. They agree that the ancestors of humans produced complex, ‘musical’ vocalisations for similar functions as songbirds do today. These songs would be sung in various contexts and they argue that general mammalian learning abilities would start, initially randomly, to associate particular songs or sub-components of songs with particular contexts. They argue that over time these sorts of associations would be strengthened and certain vocalisations could thus come to encode information about particular states of the world. There are a number of empirical questions that need to be answered to provide greater support for this hypothesis, but as it stands it does provide a plausible trajectory from song-like vocalisations to system more like language.

Models of iterated learning frequently start with random associations of non-compositional complex signals with particular meanings. The models demonstrate that through a process of cultural evolution these signals adapt to pressures such as learnability and expressivity such that component parts of the signal begin to systematically reflect particular features of the intended meaning (e.g. Kirby, 2001; Brighton, 2002). These results suggest that once we developed the ability to attribute even simple meanings to complex signals cultural evolutionary process may explain how this mapping system can come to look more like language as we observe it today.

Human language thus depends on two key pre-requisites; the ability to learn and produce complex vocal signals and the ability to infer and attribute meanings to these signals. While neither of these pre-requisites appears to be entirely unique to humans, our particular combination of the two does appear to be unique in the natural world. Once we acquired these abilities a process of iterated learning could get started and this may then explain some of the key characteristics of language.

6.3 Concluding remarks

It has been a central assertion of this thesis that we can improve our understanding of the evolution of human language by comparing features of language with features of the communication systems of other animals. Nevertheless, human language does appear to be an unprecedented phenomenon in the evolution of life. As I argued in the introduction, the uniqueness of language stems at least in part from the complex interactions between the biological evolution of the neural and physiological systems supporting language, and from the simultaneous cultural evolution of language itself. The evolution of the abilities underlying cultural transmission in language, such as vocal learning, was therefore a key transition in the evolution of language. However this was only the start of this complex adaptive process. We are now beginning to understand some of the consequences of this process for language, but much work remains to be done and I suggest that empirical and comparative studies combined with and informed by modelling work provide a promising way forward.

Bibliography

Andersson, M. (1986). Evolution of condition-dependent sex ornaments and mating preferences: Sexual selection based on viability differences. *Evolution*, 40(4):804–816.

Andersson, M. (1994). *Sexual Selection*. Princeton University Press.

Aoki, K. (1989). A sexual selection model for the evolution of imitative learning of song in polygynous birds. *The American Naturalist*, 134(4):599–612.

Ballentine, B., Hyman, J., and Nowicki, S. (2004). Vocal performance influences female response to male bird song: an experimental test. *Behavioral Ecology*, 14(1):163–168.

Boughman, J. W. (1998). Vocal learning by greater spear-nosed bats. *Proceedings of the Royal Society B: Biological Sciences*, 265:227–233.

Boyd, R. and Richerson, P. J. (1985). *Culture and the evolutionary process*. University of Chicago Press.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54.

Brighton, H., Kirby, S., and Smith, K. (2005). Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. In Tallerman, M., editor, *Language Origins: Perspectives on Evolution*, chapter 13, pages 291–309. Oxford University Press, Oxford.

Buchanan, K., Leitner, S., Spencer, K., Goldsmith, A., and Catchpole, C. (2004). Developmental stress selectively affects the song control nucleus HVC in the zebra finch. *Proceedings of the Royal Society B: Biological Sciences*, 271(1555):2381–2386.

Buchanan, K. L., Catchpole, C. K., Lewis, J. W., and Lodge, A. (1999). Song as an indicator of parasitism in the sedge warbler. *Animal Behaviour*, 57:307–315.

Buchanan, K. L., Spencer, K. A., Goldsmith, A. R., and Catchpole, C. K. (2003). Song as an honest signal of past developmental stress in the European starling (Sturnus vulgaris). *Proceedings of the Royal Society B: Biological Sciences*, 270:1149–1156.

Catchpole, C. K. and Slater, P. J. B. (1995). *Bird Song: Biological themes and variations*. Cambridge University Press, Cambridge.

Cheney, D. and Seyfarth, R. (1990). *How Monkeys See the World: Inside the Mind of Another Species*. University of Chicago Press.

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. MIT Press, Cambridge, MA.

Clarke, E., Reichard, U. H., and Zuberbuhler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS ONE*, 1:e73.

Collins, S. (2004). Vocal fighting and flirting: the functions of birdsong. In Marler, P. and Slabbekoorn, H., editors, *Nature's Music: The Science of Birdsong*, chapter 2, pages 39–79. Academic Press Inc. (London) Ltd.

Cravioto, J. and DeLicardie, E. R. (1972). Environmental correlates of severe clinical malnutrition and language development in survivors from kwashiorkor or marasmus. In PAHO, editor, *Nutrition, the Nervous System and Behavior.* PAHO Publication Number 251., Washington D.C.

Cynx, J. (1990). Experimental determination of a unit of song production in the zebra finch. *Journal of Comparative Psychology*, 104:3–10.

Darwin, C. (1879). *The Descent of Man, and Selection in Relation to Sex*. John Murray, London, 2nd edition. Reprinted in 2004 by Penguin.

de Boer, B. (2005). Cultural and biological evolution of phonemic speech. In Capcarrere, M. S., Freitas, A. A., Bentley, P. J., Johnson, C. G., and Timmis, J., editors, *Advances in Artificial Life*, pages 614–623. Springer LNCS.

Deacon, T. (1997). *The Symbolic Species*. Penguin, London.

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, Cambridge, MA.

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

Draganoiu, T. I., Nagle, L., and Kreutzer, M. (2002). Directional female preference for an exaggerated male trait in canary (Serinus canaria) song. *Proceedings of the Royal Society B: Biological Sciences*, 269:2525–2531.

Dunbar, R. (1996). *Grooming, Gossip and the Evolution of Language*. Faber and Faber, London.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., Monaco, A. P., and Paabo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 418:869–872.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, 4(7):258–267.

Fitch, W. T. (2005). The evolution of language: a comparative review. *Biology and Philosophy*, 20:193–230.

Futuyma, D. J. (2005). *Evolution*. Sinauer Associates.

Galler, J. R. and Barrett, L. R. (2001). Children and famine: long-term impact on development. *Ambulatory Child Health*, 7:85–95.

Geissmann, T. (2000). Gibbon song and human music from an evolutionary perspective. In Wallin, N. L., Merker, B., and Brown, S., editors, *The origins of music*, pages 103–123. MIT Press, Cambridge, MA.

Gold, E. M. (1978). Complexity of automaton identification from given data. *Information and Control*, 37:302–320.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, 144:517–546.

Grunwald, P. D. (2007). *The Minimum Description Length Principle*. MIT Press, Cambridge, MA.

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., and Scharff, C. (2007).

Incomplete and inaccurate vocal imitation after knockdown of Foxp2 in songbird basal ganglia nucleus area X. *PLoS Biology*, 5(12):321.

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D., and Scharff, C. (2004). Foxp2 expression in avian vocal learners and non-learners. *Journal of Neuroscience*, 24(13):3164–3175.

Hansen, P. (1979). Vocal learning: Its role in adapting sound structures to long-distance propagation and a hypothesis on its evolution. *Animal Behaviour*, 27:1270–1271.

Hasselquist, D. (1998). Polygyny in great reed warblers: A long-term study of factors contributing to male fitness. *Ecology*, 79(7):2376–2390.

Hauser, M. (2001). What's so special about speech? In Dupoux, E., editor, *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, chapter 23, pages 417–434. MIT Press, Cambridge, MA.

Hauser, M. D. (1996). *The Evolution of Communication*. MIT Press, Cambridge, MA.

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

Heywood, J. S. (1989). Sexual selection by the handicap mechanism. *Evolution*, 43(7):1387–1397.

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

Holy, T. E. and Guo, Z. (2005). Ultrasonic songs of male mice. *PLoS Biology*, 3(12):e386.

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*, 16:319–326.

Hurford, J. R. (2003). The language mosaic and its evolution. In Christiansen, M. and Kirby, S., editors, *Language Evolution*, Studies in the Evolution of Language, chapter 3, pages 38–57. Oxford University Press, Oxford.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, Oxford.

Janik, V. M. and Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior*, 26:59–99.

Jarvis, E. D. (2004a). Brains and birdsong. In Marler, P. and Slabbekoorn, H., editors, *Nature's Music: The Science of Birdsong*, chapter 8, pages 226–271. Academic Press Inc. (London) Ltd.

Jarvis, E. D. (2004b). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, 1016:749–777.

Jarvis, E. D., Ribeiro, S., da Silva, M. L., Ventura, D., Vielliardk, J., and Mello, C. V. (2000). Behaviourally driven gene expression reveals song nuclei in hummingbird brain. *Nature*, 406:628–632.

Kirby, S. (1999). *Function, selection and innateness: the emergence of language universals*. Oxford University Press, Oxford.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Knight, C., editor, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 303–323. Cambridge University Press, Cambridge.

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):102–110.

Kirby, S. (2007). The evolution of language. In Dunbar, R. and Barrett, L., editors, *Oxford Handbook of Evolutionary Psychology*, pages 669–681. 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 of the USA*, 105(31):10681–10686.

Kirkpatrick, M. (1982). Sexual selection and the evolution of female choice. *Evolution*, 36(1):1–12.

Kroodsma, D. (2004). The diversity and plasticity of birdsong. In Marler, P. and Slabbekoorn, H., editors, *Nature's Music: The science of birdsong*, chapter 4, pages 108–131. Academic Press Inc. (London) Ltd.

Lachlan, R. F. and Feldman, M. W. (2003). Evolution of cultural communication systems: the coevolution of cultural signals and genes encoding learning preferences. *Journal of Evolutionary Biology*, 16:1084–1095.

Lachlan, R. F. and Servedio, M. R. (2004). Song learning accelerates allopatric speciation. *Evolution*, 58:2049–2063.

Lachlan, R. F. and Slater, P. J. B. (1999). The maintenance of vocal learning by gene-culture interaction: the cultural trap hypothesis. *Proceedings of the Royal Society B: Biological Sciences*, 266:701–706.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., and Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413:519–523.

Lauay, C., Gerlach, N. M., Adkins-Regan, E., and Devoogd, T. J. (2004). Female zebra finches require early song exposure to prefer high-quality song as adults. *Animal Behaviour*, 68:1249–1255.

Leitao, A. and Riebel, K. (2003). Are good ornaments bad armaments? male chaffinch perception of songs with varying flourish length. *Animal Behaviour*, 66:161–167.

Leitao, A., van Dooren, T. J. M., and Riebel, K. (2004). Temporal variation in chaffinch Fringilla coelebs song: interrelations between the trill and flourish. *Journal of Avian Biology*, 35:199–203.

Livingstone, D. (2002). The evolution of dialect diversity. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*, chapter 5, pages 99–118. Springer Verlag, London.

Locke, J. L. and Bogin, B. (2005). Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, 29:259–280.

Marcus, G. F. and Fisher, S. E. (2003). Foxp2 in focus: what can genes tell us about speech and language? *Trends in Cognitive Sciences*, 7(6):257–262.

Marler, P. (2004). Innateness and the instinct to learn. *Annals of the Brazilian Academy of Sciences*, 76:189–200.

Marler, P. and Doupe, A. J. (2000). Singing in the brain. *Proceedings of the National Academy of Sciences of the United States of America*, 97(7):2965–2967.

Marler, P. and Slabbekoorn, H., editors (2004). *Nature's Music: The Science of Birdsong*. Academic Press Inc. (London) Ltd.

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

Maynard Smith, J. and Szathmáry, E. (1995). *The major transitions in evolution*. Oxford University Press, Oxford.

Merker, B. and Okanoya, K. (2007). The natural history of human language: bridging the gap without magic. In Lyon, C., Nehaniv, C., and Cangelosi, A., editors, *Emergence of communication and language*, pages 403–420. Springer.

Miller, G. (2000). *The Mating Mind*. Vintage.

Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA.

Mithen, S. (2006). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. Phoenix.

Naguib, M., Nemitz, A., and Gil, D. (2006). Maternal developmental stress reduces reproductive success of female offspring in zebra finches. *Proceedings of the Royal Society B: Biological Sciences*, 273(1596):1901–1905.

Nettle, D. (1999). *Linguistic diversity*. Oxford University Press, Oxford.

Nottebohm, F. (1972). The origins of vocal learning. *The American Naturalist*, 106:116–140.

Nowicki, S., Hasselquist, D., Bensch, S., and Peters, S. (2000). Nestling growth and song repertoire size in great reed warblers: evidence for song learning as an indicator mechanism in mate choice. *Proceedings of the Royal Society B: Biological Sciences*, 267:2419–2424.

Nowicki, S., Peters, S., and Podos, J. (1998). Song learning, early nutrition and sexual selection in songbirds. *American Zoologist*, 38:179–190.

Nowicki, S. and Searcy, W. A. (2004). Song function and the evolution of female preferences: Why birds sing, why brains matter. *Annals of the New York Academy of Sciences*, 1016(1):704–723.

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

Nowicki, S., Searcy, W. A., and Peters, S. (2002b). Quality of song learning affects female response to male bird song. *Proceedings of the Royal Society B: Biological Sciences*, 269:1949–1954.

Oberweger, K. and Goller, F. (2001). The metabolic cost of birdsong production. *Journal of Experimental Biology*, 204:3379–3388.

O'Connor, R. J. (1984). *The Growth and Development of Birds*. John Wiley & Sons.

O'Donald, P. (1980). *Genetic models of sexual selection*. Cambridge University Press, Cambridge.

Okanoya, K. (2002). Sexual display as a syntactic vehicle: The evolution of syntax in birdsong and human language through sexual selection. In Wray, A., editor, *The Transition to Language*, chapter 3, pages 46–63. Oxford University Press, Oxford.

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

Okanoya, K. and Hosino, T. (2000). Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. *NeuroReport*, 11:2091–2095.

Okanoya, K. and Merker, B. (2007). Neural substrates for string-context mutual segmentation. In Lyon, C., Nehaniv, C., and Cangelosi, A., editors, *Emergence of communication and language*, pages 421–434. Springer.

Okanoya, K., Takashima, A., Okuhara, A., Ikebuchi, M., Otsuka, R., and Wada, M. (2005). Syntactical complexity of male songs stimulates reproductive development in female Bengalese finches. *Zoological Science*. submitted for publication.

Okanoya, K. and Yamaguchi, A. (1997). Adult bengalese finches require real-time auditory feedback to produce normal song syntax. *Journal of Neurobiology*, 33(4):343–356.

Otto, S. and Day, T. (2006). *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press.

Oudeyer, P.-Y. (2005). How phonological structures can be culturally selected for learnability. *Adaptive Behavior*, 13:269–280.

Payne, R. S. and McVay, S. (1971). Songs of humpback whales. *Science*, 173:585–597.

Pinker, S. and Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13:707–784.

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

Prather, J. F., Peters, S., Nowicki, S., and Mooney, R. (2008). Precise auditoryvocal mirroring in neurons for learned vocal communication. *Nature*, 451:305–310.

Ralls, K., Fiorelli, P., and Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals: Phoca vitulina. *Canadian Journal of Zoology*, 63(5):1050–1056.

Riebel, K. (2003). The “mute” sex revisited: Vocal production and perception learning in female songbirds. *Advances in the Study of Behavior*, 33:49–86.

Ritchie, G. R. S. and Kirby, S. (2005). Selection, domestication, and the emergence of learned communication systems. In *Proceedings of AISB 2005: Social Intelligence and Interaction in Animals, Robots and Agents*.

Ritchie, G. R. S. and Kirby, S. (2006). Modelling the transition to learned communication: An initial investigation into the ecological conditions favouring cultural transmission. In Cangelosi, A., Smith, A. D. M., and Smith, K., editors, *The Evolution of Language: Proceedings of the 6th International Conference on the Evolution of Language*, pages 283–290.

Ritchie, G. R. S. 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 387–402. Springer Verlag.

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

Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192.

Rose, G. J., Goller, F., Gritton, H. J., Plamondon, S. L., Baugh, A. T., and Cooper,

B. G. (2004). Species-typical songs in whitecrowned sparrows tutored with only phrase pairs. *Nature*, 432:753–758.

Sasahara, K. and Ikegami, T. (2003). Coevolution of birdsong grammar without imitation. In *Advances in Artificial Life*, volume 2801, pages 482–490. Springer LNCS.

Sasahara, K. and Ikegami, T. (2004). Song grammars as complex sexual displays. In *Artificial Life 9*, pages 305–314. Springer LNCS.

Searcy, W. A. and Yasukawa, K. (1996). Song and female choice. In Kroodsma, D. E. and Miller, E. H., editors, *The Ecology and Evolution of Acoustic Communication in Birds*, chapter 25, pages 454–473. Cornell University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423 and 623–656.

Sipser, M. (1997). *Introduction to the Theory of Computation*. PWS Publishing Company.

Slocombe, K. E. and Zuberbuhler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology*, 15:1779–1784.

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

Soma, M., Takahasi, M., Ikebuchi, M., Yamada, H., Suzuki, M., Hasegawa, T., and Okanoya, K. (2006). Early rearing conditions affect the development of body size and song in bengalese finches. *Ethology*, 112:1071–1078.

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

Spencer, K. A., Buchanan, K. L., Goldsmith, A. R., and Catchpole, C. K. (2004). Developmental stress, social rank and song complexity in the European starling (Sturnus vulgaris). *Proceedings of the Royal Society B: Biological Sciences (Suppl.)*, 271:S121–S123.

Spencer, K. A., Wimpenny, J. H., Buchanan, K. L., Lovell, P. G., Goldsmith, A. R., and Catchpole, C. K. (2005). Developmental stress affects the attractiveness of male song and female choice in the zebra finch (Taeniopygia guttata). *Behavioral Ecology and Sociobiology*, 58:423–428.

Stromswold, K. (2001). The heritability of language: A review and metaanalysis of twin, adoption and linkage studies. *Language*, 77:647–723.

Suzuki, R., Buck, J. R., and Tyack, P. L. (2006). Information entropy of humpback whale songs. *Journal of the Acoustical Society of America*, 119:1849–1866.

Teal, T. K. and Taylor, C. E. (2000). Effects of compression on language evolution. *Artificial Life*, 6:129–143.

Teramitsu, I., Kudo, L. C., London, S. E., Geschwind, D. H., , and White, S. A. (2004). Parallel foxp1 and foxp2 expression in songbird and human brain predicts functional interaction. *Journal of Neuroscience*, 24(13):3152–3163.

Thomas, R. J. (2002). The costs of singing in nightingales. *Animal Behaviour*, 63:959–966.

Tubaro, P. L. and Segura, E. T. (1994). Dialect differences in the song of zonotrichia capensis in the southern pampas: A test of the acoustic adaptation hypothesis. *The Condor*, 96:1084–1088.

Tyack, P. L. and Sayigh, L. S. (1997). Vocal learning in cetaceans. In Snowdon, C. T. and Hausberger, M., editors, *Social influences on vocal development*, chapter 11, pages 208–233. Cambridge University Press, Cambridge.

Vallet, E., Beme, I., and Kreutzer, M. (1998). Two-note syllables in canary songs elicit high levels of sexual display. *Animal Behaviour*, 55:291–297.

Vallet, E. and Kreutzer, M. (1995). Female canaries are sexually responsive to special song phrases. *Animal Behaviour*, 49:1603–1610.

Wiles, J., Watson, J., Tonkes, B., and Deacon, T. W. (2005). Transient phenomena in learning and evolution: Genetic assimilation and genetic redistribution. *Artificial Life*, (11):177–188.

Williams, J. M. and Slater, P. J. B. (1993). Does chaffinch fringilla coelebs song vary with the habitat in which it is sung. *Ibis*, 135:202–208.

Zahavi, A. (1975). Mate selection – a selection for a handicap. *Journal of Theoretical Biology*, 53:205–214.

Appendix A

Publications

This appendix contains three papers based on work presented in this thesis which have been published as (Ritchie and Kirby, 2006), (Ritchie and Kirby, 2007) and (Ritchie et al., 2008).

MODELLING THE TRANSITION TO LEARNED COMMUNICATION: AN INITIAL INVESTIGATION INTO THE ECOLOGICAL CONDITIONS FAVOURING CULTURAL TRANSMISSION

GRAHAM R. S. RITCHIE and SIMON KIRBY

Language Evolution and Computation Research Unit,
University of Edinburgh, Edinburgh, UK
graham.ritchie@ed.ac.uk / simon@ling.ed.ac.uk

Vocal learning is a key component of the human language faculty, and is a behaviour we share with only a few other species in nature. Perhaps the most studied example of this phenomenon is bird song which displays a number of striking parallels with human language, particularly in its development. In this paper we present a simple computational model of bird song development and then use this in a model of evolution to investigate some of the ecological conditions under which vocal behaviour can become more or less reliant on cultural transmission.

1. Introduction

One of the most unusual characteristics of language, when compared to many of the other communication systems found in nature, is the extent to which it relies on vocal signals transmitted culturally rather than genetically. This is of considerable interest as other modelling work has demonstrated the role that cultural transmission, via ‘iterated learning’, may play in explaining many prominent features of human languages, e.g. the emergence of compositional syntax (e.g. Brighton, 2002), regular and irregular word forms (Kirby, 2001), and dialects (Livingstone, 2002). The evolution of learning can therefore be seen as a key transition in the evolution of human language.

Vocal learning is a comparatively rare evolutionary development, it appears to have only evolved in three groups of mammals: humans, bats and cetaceans, and three groups of birds: songbirds, hummingbirds, and parrots (Jarvis, 2004). Of these, the development of bird song and human language have a number of striking similarities, e.g. both nestlings and human babies have a critical period for learning, both rely on auditory feedback for normal development, and both exhibit a form of early babbling (known as subsong in birds) (Doupe & Kuhl, 1999). This suggests that there may be strong epigenetic constraints on the evolution of a learned vocal system (Jarvis, 2004), and so studying the evolution of learning in bird song may help us to elucidate possible ecological factors which played a role in the transition to learned communication in our own species.

In this paper we use a computational model to investigate the possible role of two very simple ecological conditions which we think may affect the transition to learning; namely the reliability of cultural transmission, and the stage of life at which communication is required.

2. The auditory template model of song development

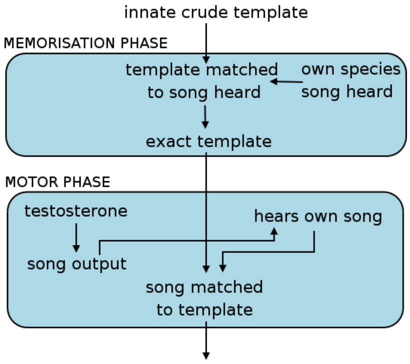


Figure 1. The auditory template model of song development. Figure after (Catchpole & Slater, 1995).

The song learning behaviour of many different species of the oscine passerines has been extensively studied, for an introduction see the reviews in (Catchpole & Slater, 1995) and (Marler & Slabbekoorn, 2004). The exact pattern of song development varies greatly among different species, but in attempting to capture the general features, bird song biologists have developed what is known as the 'auditory template model' of song learning, depicted in figure 1. This model posits two distinct phases to song learning; an early memorisation phase in which songs that are heard as an infant which are recognised as conspecific by an innate 'crude template' are memorised, and a later motor phase when song production is trained to produce songs that match the learned template. This behaviour can be contrasted with the sub-oscine passerines which appear to have a largely innately specified song, and will develop normal song production without hearing conspecific song and without auditory feedback.

3. A simple computational model

We take this model as our inspiration and develop a computational model of the two stages of learning in bird song, described in the proceeding two sections. We then use this model to investigate some conditions under which song perception and production can come to be increasingly influenced by cultural transmission.

3.1. *Phase 1: Observational learning*

To model the memorisation phase of song learning we hypothesise a module which we term the Species Recognition Device (SRD). This is intended to model the auditory biases birds appear to show towards conspecific song.

We model the SRD as a note transition matrix which defines the transition probabilities between every available note (or song element)[a]. We assume that the notes are fixed and identical for every agent in the simulation, and the number of notes used here was 6. We realise that this is unrealistic and that many species learn the form of song elements from their tutors as well as the element sequence. We also realise that element transitions or sequence are not the only cues birds use to identify conspecific song. However, a note transition matrix provides us with a simple and computationally tractable model of these sorts of biases.

Each agent in the model has 'genes' which code for an innate SRD, this is intended to model the 'crude template' as described above. An agent uses its SRD to categorise songs it hears as either conspecific or not by comparing the note transitions in the song with the transition probabilities in the matrix.

Such a matrix can be more or less biased to a particular song-type, if all the probabilities in the matrix are equal then the matrix has no preference to any particular song, while if each row has exactly one high probability transition, the matrix is maximally biased to one particular song. We can measure this bias by calculating the Shannon entropy for each transition distribution, and we can measure the preference of a matrix for a particular song by comparing the transitions found in the song and the probabilities in the matrix. We have used these measures of matrix preference and bias in earlier work (Ritchie & Kirby, 2005), and the reader is referred there for a more detailed definition.

An agent's adult SRD is also subject to being altered by songs heard in early life, we model this by 'exposing' each agent to 100 songs from its environment and getting it to select the ones preferred by its innate SRD (crude template). The note transitions in the songs that are selected at this stage are then reinforced in the agent's SRD to produce the agent's adult SRD, or 'exact template'. The degree to which an agent's SRD is modifiable by songs heard in early life is determined by genes which code for the agents SRD plasticity (SRDP), this will be a value between 0 and 1, with 0 meaning the innate SRD is entirely fixed, and 1 meaning that the agent relies only on songs heard early in life to construct its adult SRD.

3.2. *Phase 2: Reinforcement learning*

The SRD as described in the previous section models an agent's sensory biases (or lack thereof) to a particular song-type. We also require a model of song production. We also model this as a note transition matrix[b], but here the probabilities determine the probabilities of singing one note after another. We call this the Song Production Device (SPD). Just as for the SRD, an agent encodes innate biases for

---

[a] While we implement the SRD as a note transition matrix here, we hope that this component could be modelled in many different ways, e.g. as a neural net with the initial weights specified genetically.

[b] Again, we hope that the SPD component could be modelled in a number of different ways, not necessarily using the same mechanism as for the SRD.

its SPD in its genes.

To model plasticity in the production mechanism, we allow the SPD to be trained by reinforcement learning using the agent's SRD as a critic, using a very simple learning algorithm. This is intended to model the process by which a bird uses its memorised exact template to guide its vocal development. As for the SRD, the degree to which the adult SPD is allowed to be influenced by learning, the SPD plasticity, (SPDP), is determined genetically. If the plasticity is 0 then the SPD is not influenced at all by the learning procedure described below, higher values mean the SPD becomes increasingly influenced by learning.

The SPD is trained by getting the agent to produce a song and then to 'listen' to this song with its adult SRD, if a note transition in the song is 'accepted', i.e. has a high probability in the SRD matrix, that transition's probability is increased slightly in the SPD. This process is repeated 250 times, after which the agent's SPD is said to have 'crystallised' and will not change again in he agent's lifetime[c].

3.3. *Determining fitness*

We define an agent's fitness as its ability to recognise and be recognised by conspecifics. This seems a reasonable model of one of the main pressures acting on song (Catchpole & Slater, 1995), although there are of course many other pressures acting on song in he wild (e.g. sexual selection for variation, adaptation to the local acoustics etc.), and we hope to model some of these in future work.

To calculate an agent's fitness we perform 250 fitness trials. In each trial we get the agent to produce a song using its crystallised SPD and we then randomly select another member of the population and check that this second agent correctly recognises the song using its adult SRD. We also get the second agent to produce a song and check that the first agent correctly recognises the song. Every correct recognition means that the agent's fitness is incremented by 1.

Defining fitness in this way means that there is a strong selection pressure for the agents to develop and maintain a stereotypical and easily recognised species-specific song. As the SRD is modelled as note transition probability matrix, this corresponds to a matrix with a single high probability transition for each individual note. In short, in this environment it is adaptive to have strongly biased matrices.

3.4. *Overall model design*

The overall model works with an evolving population of 100 agents. As we want to investigate how a genetically specified song can come to be learned we initialise the agents innate SPD and SRD genes to one particular song "abcd", and the plasticity genes to 0. This means that the population will start off receiving maximal

---

[c] Unfortunately we do not have space to describe the learning algorithm in detail here. Further details are available upon request, and will be described more fully in future work.

fitness values and any mutations that degrade an agent's ability to sing and recognise conspecific song will be selected against. Each agent in each generation then goes through the following 'life stages':

Birth The agent's innate SRD and SPD, along with its SRDP and SPDP, are decoded from its genes.

Development Each agent is exposed to the songs of the previous generation, and picks those which will be used for learning using its innate SRD. The agent then goes through the two stages of learning described above to give them their adult SRD and crystallised SPD.

Adulthood The agents are tested in 250 fitness trials as described above to see how many times it can correctly recognise a bird of its own species and how many times its song is correctly recognised by a bird of its own species. These values are summed to give a bird's fitness score.

Reproduction Parents from the population are selected probabilistically according to their fitness score and their genes are recombined and subject to a low mutation rate to produce new child agents[d].

Death Each bird in the population is sampled 5 times and the resulting songs are stored for the next generation to learn from. All of the current birds in the population are removed and their children become the new population.

We repeat this process over many generations and record various measures over the course of a run.

4. Experiments

In this initial investigation we only model two very simple ecological conditions:

Environmental reliability For the first experiment we vary the reliability of the environment, that is the degree to which the previous generation's songs are faithfully recorded and then passed on to the new generation to learn from. We have two conditions: a reliable environment where we keep 80% of the previous generation's songs, and an unreliable environment where we keep only 20% of the previous generation's songs. The remaining songs are randomly generated songs which use the same notes and are constrained to within the same length as the agents' songs. This intended to model heterospecific song or other extraneous sounds in the birds environment.

Timing of song requirement In the first experiment we only test the bird's fitness after learning has taken place, in this experiment we also check the bird's fitness *before* learning. This intended to model a possible environment in which song is

---

[d] This is implemented with a standard genetic algorithm (GA), using tournament selection, a crossover rate of 0.7 and a mutation rate of 0.01. Mutation is modelled by simply replacing the gene that is to be mutated with a uniform random number between 0 and 1.

required immediately after birth as well as later in life.

5. Results

We provide results for each of the three different conditions described above in figure 2. The measures shown in each are the population average fitness, SPDP, SRDP, SPD change and SRD change. The SPD and SRD change are simply the absolute difference of the bias value of the innate and adult matrices (as discussed in section 3.1 above). We measure this as well as the plasticity values as the plasticity values can vary *without* a correlated variation in the change values (as demonstrated in figure 2c).

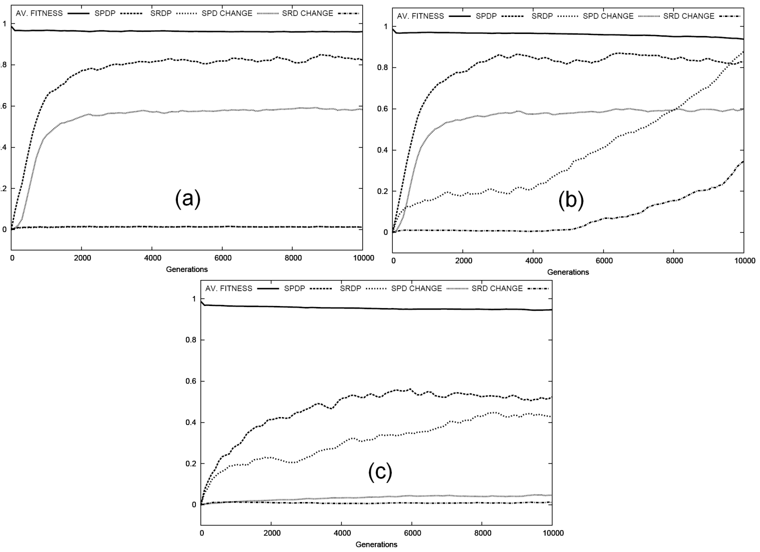


Figure 2. Results for the three different environments. The X-axis in each graph is the number of generations, set to 10000 for all results shown here. The Y-axis in each graph measures the population average fitness, SPDP, SRDP, SPD change and SRD change for each different condition. Graph (a) shows results for an unreliable environment where only 20% of the previous generation's song are faithfully passed on. Graph (b) depicts a reliable environment where 80% of the songs are passed on. Graph (c) shows results for a reliable environment in which the agents' fitness is checked both before and after learning. These results are the averages of 10 separate runs for each condition with a different random number generator seed for each. We have smoothed the graphs to allow us to better see the overall trends.

In all of the conditions we found that fitness stayed fairly fixed throughout all of the runs. However, the degree to which song remained being transmitted genetically depended on the environment, as demonstrated by the different values of SRDP and SRD change at the end of each simulation.

In the unreliable environment the population cannot count on hearing conspecific song as infants. The agents therefore have to keep transmitting their song genetically, as demonstrated by the much lower SRD change and SRDP at the end of the run in figure 2a. In contrast, in the reliable environment shown in figure 2b, towards the end of the runs the population begins to transmit their song culturally as demonstrated by the coincident rise in the population's SRDP and SRD change.

In both experiments, however, the SPD change and SPDP quickly rise, indicating that the SPD is always being trained using the adult SRDs and the reliability of the environment appears to have no bearing on this. As long as the adults can construct a faithful copy of the species song in their SRDs as a result of either cultural or genetic transmission, it can always be used to train the SPD, and so there is no pressure for the copy of song stored in the SPD to be transmitted genetically and mutation pressure quickly erodes the genetic copy.

Figure 2c show results when the timing of song requirement is changed, where we test an agent's fitness both before and after learning. The SPD and SRD change values stay low throughout the run, demonstrating that SPD and SRD copies of song remain genetically transmitted throughout the run. The average SPDP and SRDP values drift to around 0.5 as there is no selection pressure acting to maintain these at any particular value.

6. Discussion

The results described here predict two simple environmental conditions which could affect the transition to a learned communication system; the reliability of the cultural environment, and stage of life at which communicative behaviour is required. These conditions seem fairly widely applicable and it seems reasonable that these conditions may have played a role in the transition to increased reliance on learning in human communication as well.

We think that this model also provides an interesting case study of the interaction of genetic and cultural transmission and phenotypic plasticity. We see that where the environment is reliable enough, and 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. Cultural transmission can thus, in some conditions, be seen as a *masking* force (Deacon, 2003) on genetic transmission, with a similar end result to that we found in earlier work (Ritchie & Kirby, 2005) for rather different environmental conditions.

Another interesting result is that in all of the experiments described here the agents come to rely solely on their auditory copy of song (in the SRD) to guide later production behaviour. We feel that this again represents a form of genetic parsimony, as it seems rather inefficient for an agent to store two 'copies' of their song genetically, even though these copies are likely to be represented in rather different ways; one being a sensory and the other a motor mechanism. Nevertheless, if there is enough phenotypic plasticity to allow these to interact, and if the genetic 'cost' of this plasticity is lower than the cost of encoding a song genetically, we see that even in the unreliable environment the agents rely on only their auditory copy, but need it always be this way round? In the case of bird song it seems so, as a bird only needs to produce a song when it is sexually mature while it needs to be able recognise conspecific song earlier. This means that the song recognition system should be more genetically constrained than the song production system, which seems to match the biological data.

While this may be true of bird song it is not so clear for human language as children become capable talkers well before puberty. In future work we would like to relax some of the assumptions built into the current model with regard to the timing of each of the learning phases and allow this to be under genetic control. The two ecological conditions we discuss here are the simplest relevant condition we could think of, and we would also like to model other relevant ecological conditions, such as sexual selection pressure, to see what role these may play in conjunction with the conditions investigated here.

References

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*(1), 25–54.

Catchpole, C. K., & Slater, P. J. B. (1995). *Bird song: Biological themes and variations.* Cambridge University Press.

Deacon, T. (2003). Multilevel selection in a complex adaptive system: the problem of language origins. In B. Weber & D. Depew (Eds.), *Evolution and learning: the baldwin effect reconsidered* (pp. 81–106). MIT Press, Cambridge, MA.

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

Jarvis, E. D. (2004). Brains and birdsong. In P. Marler & H. Slabbekoorn (Eds.), *Nature's music: The science of birdsong* (pp. 226–271). Academic Press Inc. (London) Ltd.

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)*, 102–110.

Livingstone, D. (2002). The evolution of dialect diversity. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 99–118). London: Springer Verlag.

Marler, P., & Slabbekoorn, H. (2004). *Nature's music: The science of birdsong.* Academic Press Inc. (London) Ltd.

Ritchie, G., & Kirby, S. (2005). Selection, domestication, and the emergence of learned communication systems. In *Proceedings of AISB 2005: Social intelligence and interaction in animals, robots and agents.*

A possible role for selective masking in the evolution of complex, learned communication systems

Graham R. S. Ritchie and Simon Kirby

Language Evolution and Computation Research Unit
University of Edinburgh

Abstract. One of the most distinctive characteristics of human language is the extent to which it relies on learned vocal signals. Communication systems are ubiquitous in the natural world but vocal learning is a comparatively rare evolutionary development [Jarvis, 2004]. In this paper we take one example of this phenomena, bird song, which displays some remarkable parallels with human language [Doupe and Kuhl, 1999], and we focus on one particular case study, that of the Bengalese finch (*Lonchura striata* var. *domestica*), a domesticated species whose song behaviour differs strikingly from its feral ancestor in that it has complex syntax and is heavily influenced by early learning [Okanoya, 2002]. We present a computational model of the evolutionary history of the Bengalese finch which demonstrates how an increase in song complexity (in some sense) and increased influence from early learning could evolve spontaneously as a result of domestication acting to mask the natural selection pressure on song behaviour. We argue that this may provide an insight into how increased reliance on vocal learning could evolve in other communication systems, including human language.

1 Introduction

The human capacity for language is one of our most distinctive characteristics. While communication systems abound in the natural world, human language distinguishes itself in terms of its communicative power, flexibility and complexity. One of the most unusual features of human language, when compared to the communication systems of other species, is the degree to which it involves learning. Just how much of language is innate and how much is learned is an ongoing controversy, but it is undeniable that the specific details of any particular language must be learned anew every generation. We do, of course, bring a great deal of innate resources to bear on our language learning process, and the results these innate biases have on the development of languages may explain a great deal about the structure of the languages we see today. But still every child in every new generation must go through a lengthy process of language acquisition if they are to become normal language users.

Once in place, this inter-generational process of language acquisition and use, or *iterated learning* [Kirby and Hurford, 2002], can give rise to cultural evolution, which studies have shown may explain many prominent phenomena of human language, including the emergence of dialects and, by extension, separate languages [Livingstone, 2002], regular and irregular word forms [Kirby, 2001] and compositional syntax, e.g. [Brighton, 2002].

The emergence of learning can therefore be seen as a major transition in the evolution of language and we would like to better understand the evolutionary pressures and factors which caused this transition. A natural point at which to start such an investigation is to look at the communication systems of other animals to see if there are any parallels which might illuminate the relevant ecological factors. Much comparative research has been carried out with the non-human primates, but despite some fascinating results, it seems that their natural communication systems are very different to language, including the fact that learning plays a much less prominent role. In fact it appears that vocal learning systems have evolved in only three groups of mammals: humans, bats and cetaceans, and three groups of birds: songbirds, hummingbirds, and parrots [Jarvis, 2004] (though there is initial evidence of vocal learning in other species, including some elephants and seals.)

In this paper we concentrate on bird song as it has many striking parallels with language, particularly the way in which it is learned, as Darwin noted in *The Descent of Man*:

> The sounds uttered by birds offer in several respects the nearest analogy to language, for all the members of the same species utter the same instinctive cries expressive of their emotions; and all the kinds that sing, exert their power instinctively; but the actual song, and even the call notes, are learnt from their parents or foster-parents. [Darwin, 1879]

Since Darwin's day much research has been carried out into bird song and, to take Tinbergen's four perspectives of ethology, we now know a great deal about its mechanism, development, function and evolution. However, despite much research, in general the evolutionary function of song learning remains unclear [Slater, 2003]. The parallels between bird song and human language have also been further elaborated as modern techniques have allowed us to establish the neural mechanisms of both song and language [Doupe and Kuhl, 1999].

2 A case study

Recent studies by Kazuo Okanoya of a domesticated species of finch, the Bengalese finch (*Lonchura striata* var. *domestica*), and its feral ancestor, the white-backed munia (*Lonchura striata*) which still lives in the wild throughout Asia, provide an interesting case study of the interaction of learning and evolution in bird song. The Bengalese finch sings a song with complex[1] finite state syntax which is heavily influenced by early auditory

---

[1] Okanoya defines song complexity as the song linearity, i.e. the total number of unique song notes divided by the number of unique note-to-note transitions. We are not

experience. Surprisingly, the munia sings a strikingly simpler, more linear song which is less influenced by early learning. In other words, in a relatively short period of domestication, there have been radical changes in song behaviour. This has happened even though the domesticated species has been artificially bred for plumage rather than song.

[Okanoya, 2004] has identified the neural mechanism underlying this difference in behaviour and has shown that while Bengalese chicks are able to learn the songs of munia tutors, munia chicks are not able to learn all aspects of the more complex Bengalese song, demonstrating that there is a physiological, as opposed to cultural, basis for this difference.

2.1 Okanoya's hypothesis

As experiments have shown that both female munias and female Bengalese finches prefer the more complex song, [Okanoya, 2002] argues that it is sexual selection which drove the increase in complexity. He argues that domestication freed the Bengalese finch from the pressure of predation and other pressures associated with life in the wild which had previously held song complexity in check. According to Okanoya, the more complex song of the Bengalese finch may therefore be seen as an honest signal of fitness [Zahavi, 1975]; a fitter bird can afford a more complex song. [Sasahara and Ikegami, 2004] show with a computational model of the finch data that, under some assumptions about female preferences and perception, song complexity could indeed increase as a result of sexual selection.

2.2 Deacon's hypothesis

Reviewing the same data, Deacon (p.c.) agrees that domestication masked the natural selection pressure keeping the munia's song simple, but argues that the increase in complexity happened *without* direct selection on the trait. Essentially, he posits that domestication shielded the trait from selection which allowed random genetic drift to erode innate song biases in the munia. This allows previously minor influences, such as mnemonic biases and early auditory experience, 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-differentiated*. Deacon goes on to argue that this process of masking and subsequent de-differentiation is a potential explanation for the evolution of complex functional synergies such as the neural mechanisms for song production now present in the Bengalese finch, and, he argues, in the human capacity for language.

The concept of selective masking and its effect on the evolution of language are explored in more detail in [Deacon, 2003]. [Wiles et al., 2005] demonstrate with a computational model how this kind of masking (and

---

entirely satisfied with this as a measure of complexity, as discussed in section 4.2, but we use the term in Okanoya's sense throughout this paper.

later unmasking) effect may have played a role in the functional integration of groups of genes underlying complex traits, e.g. the mammalian colour vision system.

3 A computational model

In order to evaluate Deacon's hypothesis and to try to establish if such behaviour could evolve spontaneously as a result of domestication, we have developed a computational model of the finch data. The model is designed to be reasonably biologically plausible, and also general enough that it could be extended to other species. The model works with an evolving population of agents, or birds, and the main stages in the simulation are listed here, details of each stage are given below:

Birth The bird's song filter is built up from its genotype as described in section 3.1.

Development The bird is exposed to e songs from its environment, and, using its filter, selects t songs from which it will learn (its training set) as described in section 3.2. The bird then uses the learning algorithm described in section 3.3 to learn the song grammar it will use to sing throughout its life.

Adulthood The bird is tested in f fitness trials, as described in section 3.4 to see how many times, using its filter, is can correctly recognise a bird of its own species and how many times it is correctly recognised by a bird of its own species. These values are added to give a bird's fitness score.

Reproduction Parents of the same species are selected probabilistically according to their fitness score and their chromosomes are crossed over using one-point crossover with probability pCO (set to 0.7 for all results provided here), to give a new child. Individual genes are mutated with probability $pMut$ (set to 0.05 for all results provided here, lower values have qualitatively similar results but the simulations take considerably longer to show the same effects). The mutation operator used is the 'Reflect' operator described in [Bullock, 1999].

Death Each bird in the population is sampled s times and the resulting songs are stored for the next generation to learn from. All of the current birds in the population are removed and their children become the new population.

3.1 The song filter

A bird is modelled as having a genetically coded note[2] transition matrix, which specifies a transition probability from each note to every other note

---

[2] It should be noted that while we use the term 'note' throughout this chapter, this is not intended to refer to a particular acoustic note, rather we simply use it to denote an atomic song element that can be reliably differentiated from other elements which appear in the song.

used in the simulation, including a probability for the first and final notes. The total number of notes is a parameter of the simulation, $numNotes$, but in all results provided here this was set to 8, i.e. the notes from a alphabetically through to h, this value was chosen as it appears to be the number of unique notes identifiable in both the Bengalese finch and munia's songs [Okanoya, 2002]. The matrix is coded for by a chromosome which has one real valued locus for each entry in the matrix which can vary between 0 and 1. This chromosome will thus have $(numNotes+1)^2$ loci, the 1 is added to include the transitions at the beginning and end of the song. To construct a matrix from the chromosome we take each $numNotes+1$ loci of the chromosome in turn, and normalise the values to give a probability distribution for each row of the matrix (i.e. we sum the values of each $numNotes+1$ loci and then divide each value by this sum to give a probability). An example matrix, and the chromosome that codes for it is shown in table 3.1.

	a	b	c	E
S	0.08	0.15	0.62	0.15
a	0.11	0.89	0.00	0.00
b	0.05	0.10	0.40	0.45
c	0.82	0.09	0.00	0.09

0.1	0.2	0.8	0.2	0.1	0.8	0.0	0.0	0.1	0.2	0.8	0.9	0.9	0.1	0.0	0.1

Table 1. An example note transition matrix and the chromosome that codes for it. *numNotes* here is set to 3 meaning that the chromosmoe will have $(3+1)^2 = 16$ loci (we used 8 notes in our simulations, and hence chromosomes with 81 loci). The S indicates the start of the song, and the E indicates the end of the song.

The transition matrix serves one main purpose; to establish the probability that a given song is one of the bird's own species song. This is done by establishing the average probability of each note transition in the song, as shown in equation 1 which defines the preference a given matrix m_x has for a particular song s_y, in this equation n is the number of note transitions in s_y and $m_x(t_i)$ is the entry in m_x for the ith transition of s_y. For example the preference value the matrix in table 3.1 gives for the song *cab*, which has the transitions S-c, c-a, a-b and b-E, is $\frac{0.62+0.82+0.89+0.45}{4} = 0.695$, while the preference for the song *acb* is 0.043. Note that we always include the transition to the first note and from the last note, so the empty song ‘’ has a single transition S-E, for which this matrix has a preference value of 0.15.

$$preference(m_x, s_y) = \frac{\sum_{i=0}^{n} m_x(t_i)}{n} \qquad (1)$$

The matrix can be thought of as a song ‘filter’. A song with a high probability will be more likely to pass though the filter than one with a lower probability, in our example *cab* would be much more likely to pass through the filter than *acb*. If the matrix has a single high probability transition for each note this can be thought of as a strong filter, as it will only accept songs which contain these transitions. If the matrix has even probabilities for each transition it is considered a weak filter as it accepts all songs equally.

We can measure the strength of the filter explicitly by calculating the entropy for each transition distribution (i.e. each row in the matrix), using [Shannon, 1948]'s measure. This will result in a value which ranges from 0 to $log(nValues)$, where $nValues$ is the number of probabilities in row r_x (i.e. the number of columns in the matrix). We then normalise this value into the range 0 to 1, as shown in in equation 2, which defines the normalised entropy for a given row r_x, in this equation p_i is the probability of the ith transition in r_x. The overall strength of a matrix m_x is then calculated as the average entropy of each row r in the matrix, as shown in equation 3. A filter strength of 0 means that the filter will only accept one song while a strength of 1 means that the filter will accept all songs equally. As an example, the matrix in table 3.1 has a strength value of 0.56.

$$entropy(r_x) = \frac{-\sum_{i=1}^{nValues} p_i \; log\ (p_i)}{log(nValues)} \tag{2}$$

$$strength(m_x) = \frac{\sum_{i=1}^{nRows} entropy(r_i)}{nRows} \tag{3}$$

This filter is intended to model the preferences many songbirds have for their species specific song [Catchpole and Slater, 1995]. In the model a bird uses its filter for two purposes:

1. To select its training set (the songs it will later use to learn from) from the songs it is exposed to during infancy.
2. To judge whether another bird is a member of the same species for mating or territorial defense.

In this respect, this model is similar to those used in Lachlan's models of the 'cultural trap' in bird song [Lachlan and Slater, 1999,Lachlan and Feldman, 2003]. This seems a reasonably plausible assumption, as it is known that some songbirds do have an innate preference for conspecific song both when learning songs as a nestling and also for later mate selection [Catchpole and Slater, 1995].

3.2 Selecting the training set

The infant bird is exposed to e environmental songs to select its t training songs from, both e and t are parameters of the simulation, but were set to 50 and 5 respectively for all results provided here. 5 seems a rather low value of t, but the learning algorithm is very computationally intensive and so a low value is used to speed up the simulation. The e environmental songs are randomly selected from the songs sampled from the previous generation, to compose this set each bird is sampled s times, another parameter which is set to 5 here, so for a population size $popSize$ of 100, as used here, this will contain 500 songs.

The infant bird is exposed to each of the e songs in turn and uses its filter to compute the probability it will be accepted. During experimental runs it was determined that checking that the song is accepted once did not impose enough of a pressure for the bird to correctly select conspecific song and so a song is only added if it is accepted by the filter twice successively. If the bird has not picked t songs after being exposed to all e songs, the process is repeated until t songs have been selected. The training songs are then fed into the learning algorithm described below.

3.3 Song learning

Song learning is modelled as minimum description length (MDL) induction of a probabilistic finite state machine (PFSM), closely following the algorithm described in [Teal and Taylor, 2000]. Induction of finite-state machines was chosen to model learning as [Okanoya, 2002] argues that the songs of both munias and Bengalese finches can be usefully described by a finite-state syntax. The algorithm works by firstly establishing the maximal PFSM that explicitly represents each song in the training set, the prefix tree. The algorithm then searches for nodes which can be merged which will reduce the MDL of the overall machine, whilst also ensuring that the PFSM remains deterministic. The MDL measure takes into account the amount of information (measured by the number of bits) required to code for the machine itself, and also to code for each of the training songs in terms of the machine. Essentially the algorithm searches for the most parsimonious machine in terms of the data. This approach allows a bird to generalise from its training set, whilst also always being able to reproduce each of the songs it learned from. The reader is referred to [Teal and Taylor, 2000] for a more detailed description of the algorithm used. The only difference between Teal and Taylor's and our approach is that we also take into account the probability of each note transition, given the probabilities of each transition in the training set.

3.4 Calculating a bird's fitness

To establish a bird's fitness we want to check both that its filter allows it to correctly identify its own species, and that its song is correctly identified by other birds of its species. This seems a reasonable model of the pressures acting on song in the wild [Catchpole and Slater, 1995].

To calculate an individual bird b_i's fitness we perform f fitness trials, a parameter set to 250 for the results provided here. In each fitness trial we get b_i to produce a song and we then randomly select another member of the population, b_j and check that b_j correctly recognises the song using its filter. We also get b_j to produce a song and check that b_i correctly recognises the song with its filter. Every correct recognition means that b_i's fitness is incremented by 1. With f set to 250, this means that the maximum fitness achievable is 500, or generally $2f$.

3.5 Modelling the finch data

This is a fairly general model of bird song, and so we need to set it up to match the data available on the Bengalese finch and the munia as closely as possible. The simulation passes through 3 main phases, each of which runs for 500 generations. The phases are described below.

Phase 1 We know that the white-backed munia has a very stereotyped song and that it seems to only be able to learn songs that match its species-specific song fairly closely (a munia cross-fostered with Bengalese parents is not able to learn all aspects of its tutor's song). In our model this corresponds to the munia having a strong filter. To simulate this state we seed the environmental songs with a single song type, e.g. $abcdef$. We then run the simulation for 500 generations using the fitness function and learning algorithm described above. As the environment songs are entirely identical the songs that any bird will learn from are always the same, and so they will always induce the same PFSM. This is not meant to be biologically plausible, we simply want the population to develop strong filters for a particular simple song type.

Phase 2 At the end of phase 1 we have a population of birds who sing a stereotypical song and produce offspring with a strong genetic bias to learn that song. To test if the filter can indeed help young birds recognise the appropriate song to learn from in the second phase of the run we start introducing random songs into the bird's environment, this is intended to model hetero-specific song in the environment. We model this by replacing 10% of the s sampled songs with randomly generated songs which use the same notes as the current population and which are constrained to within the same length. We realise that hetero-specific songs are unlikely to be truly 'random' in a real environment and so this may be an unrealistic modelling decision, but we simply want to model some degree of noise in the acoustic environment which the population should be able to filter out. This seems reasonable to us as if a population of birds really had a completely reliable set of songs to learn from every generation we would expect them to have lost any bias to conspecific song as this would be unnecessary, every song they heard would be conspecific, but this doesn't match the biological data. An alternative strategy that we have experimented with (but not used here) is to randomly 'mutate' notes from the songs sampled at the end of each generation with some low probability, and tests show that this produces qualitatively similar results to those provided here, but runs take much longer to show the same effects.

Phase 3 We model domestication of the population simply by ceasing to calculate fitness, but we continue to perform the crossover and mutation operations. The seems a reasonable model of domestication, as in captivity the birds no longer have to recognise their own species to successfully mate or defend their territory as the mating is now controlled by humans and they are kept in aviaries. Domestication can thus be seen to *mask* the selection pressure on these functions. We continue to introduce 10% of random songs into the environment each generation, as it seems a reasonable asumption that the birds will still be exposed to hetero-specific song, or at least other extraneous sounds, in captivity. Experimental results of this setup are described in the next section.

4 Results

The graphs in figure 1 shows several measures taken over the course of each of the three phases described above.

The first graph shows the change in the average population fitness plotted against the change in filter strengths through the three phases. Fitness values are not calculated for the poulation in phase 3. The filter strength is calculated as described in equation 3.
The second graph shows our various complexity measures plotted against each other over the entire run, the first of these the average grammar encoding length (GEL) of the population's PFSMs, this is a measure of the size, in bits, it would take to encode a PFSM using the measure defined in [Teal and Taylor, 2000]. The second measure is the average song linearity of the population's songs (Okanoya's definition of complexity), defined as the number of unique notes in each song divided by the number of unique note to note transitions. The final measure is the average linearity of the population's PFSMs calculated simply as the number of states divided by the number of transitions. A completely linear PFSM would thus have a linearity of 1, while a maximally non-linear PFSM would have a linearity equal to 1 over the number of transitions in the PFSM.

Two example PFSMs taken from the population at the end of phase 2 are shown in figure 2, and two PFSMs from the end of phase 3 are shown figure 3. The GEL and PFSM linearity values for each machine is also given.

4.1 Analysis

The results in the first graph demonstrate that the strong filters built up in phase 1, as shown by the increase in filter strength[3], enable the birds to filter out the hetero-specific songs introduced in phase 2 without any fitness decrease. We see that all 5 measures stay roughly the same throughout this phase, indicating that this is a fairly stable state. When we 'domesticate' the population in phase 3 we see a significant change in its behaviour. Immediately we see that the filters begin to weaken, and we see that the average GEL rises steadily throughout the phase

---

[3] Recall that the strongest filter would give a value of 0, and the weakest 1

indicating that the population's PFSMs are getting larger and the birds have a more varied song repertoire. At the same time we see both the song, and underlying PFSM linearity drop, indicating that the songs a bird will sing have comparatively more varied note transitions, i.e. a more 'complex' song in Okanoya's sense.

This behaviour seems to be a result of the fact that the strength of the population's filters is no longer being selectively maintained, that it they have been *masked* from selection. This allows mutations to accumulate and for the filters to become steadily weaker. This allows some of the hetero-specific songs to pass though the filter when a bird is selecting its training set, which results in the bird inducing a more varied PFSM. Essentially the domesticated populations is able to learn from much more varied sources and so early auditory experience has much more of an effect on adult song behaviour.

These results are comparable to the masking phase of [Wiles et al., 2005]. Their model, however, goes further than ours and shows that if the selection pressure were later *un*masked this could result in a selection pressure for any other abilities able to help the organism survive in the new environment (e.g. colour vision). It would be interesting to investigate what might happen to the Bengalese finch's song if its environment changed substantially again, e.g. a population was released back into the wild.

4.2 Song complexity?

[Okanoya, 2002] argues that the Bengalese finch has a much more 'complex' song than the munia. As mentioned earlier, his measure of complexity is the song linearity. He finds that the average song linearity of the munia is around 0.8 while the Bengalese finch song has a value of around 0.4. We provide results for this measure over the course of our simulations in the graph above, but on average we also see a higher value, around 0.95, for the ancestral population and a lower value, around 0.6, for the domesticated population at the end of our run.

While this measure seems a reasonably intuitive measure of song complexity (the more varied a song is, the more complex it is), it should be noted that this measure will classify an entirely random song as maximally complex. We do not want to equate randomness with complexity, but we find it hard to define a measure that can differentiate between the two. Any standard measure of the information content of a song will not be able to do so; a random song is maximally informative in information-theoretic terms. However we consider that two measures, the GEL of a bird's PFSM taken together with the linearity of the PFSM provide a reasonable estimate of the complexity of a song. A PFSM with a very small GEL and a low linearity is likely to produce more random songs, as it approaches a one state PFSM with multiple transitions back to the same state. A PFSM with a large GEL, but a very high linearity (as we see in the ancestral population in the model) will produce an entirely linear song. A PFSM with a large GEL and a relatively low linearity will produce songs that we are more happy to refer to as complex, as the GEL indicates that it has many states, and so different notes will be used in different contexts, but each state also has several transitions which means that different transitions can be made from each context. Our results demonstrate that the domesticated population does have a higher GEL and a lower PFSM linearity than the wild population and so we are tentatively happy to agree that domestication has caused an increase in song complexity. However, we are still working on developing a more satisfactory measure of song complexity.

4.3 Comparison with the biological data

Comparing these results with the data available for the Bengalese finch we find that the model does seem to capture some of the phenomena involved. Okanoya has shown that a munia chick which is not exposed to conspecific song will not sing a normal song, which seems to fit with the model. He has also shown that while Bengalese chicks can readily learn munia songs, munia chicks cannot learn the more complex Bengalese songs. In the model this difference is attributable to their different filters. The difference in the values for the song linearity in the ancestral and domesticated populations also seem to match fairly well.

As it stands though, the model does not explain why the female munia prefers the Bengalese song. We would argue that a bias for complex song may have been latent in the munia, and the fact that the munia females prefer the more complex song does not prove that this was the driving force for the change in song behaviour, although introducing such a preference into the model may help to tease these pressures apart. [Okanoya, 2004] demonstrates that the NIf region of the Bengalese finch's brain is necessary for it to be able to sing the more complex song; when surgically lesioned a Bengalese finch with previously complex song will sing a simpler, more munia-like song. We would argue that the model remains neutral to this datum, as it is possible that the munia does have this pathway present in its brain but, because it only ever learns a simpler song, does not use it.

5 Discussion

Our results demonstrate that, as Deacon initially proposed, an increase in song complexity (in some sense) and increased influence from early learning can arise *without* direct selection on either trait, simply through the process of domestication, but what is the significance of this result for the study of human language? Can studying the evolution of learning and complexity in bird song inform our study of the origins of complex language in our species? We believe that understanding the mechanisms behind the emergence of the Bengalese finch's song, and indeed the evolution of bird song in general, is valuable for evolutionary linguistics in two ways.

Firstly, it has been argued that iterated learning is a key mechanism for the origins of syntax in human language [Kirby and Hurford, 2002].

It is striking that human language differs from most other communication systems both in being transmitted through iterated learning and in having complex syntactic structure. We say "most" here but not "all". We appear to be in a very exclusive club with songbirds possibly as another member. Of course, there are important differences between iterated learning in humans and birds. For example, in the former a central constraint on transmitted languages is that they be *expressive*, in that strings must convey complex information. Bird song does not seem to carry 'meaning' in the same way, although a diversity of songs may play a role as a sexual display [Catchpole and Slater, 1995]. Nevertheless the co-occurrence of iterated learning and signal complexity in both songbirds and humans combined with the rarity of either anywhere else in nature cannot be ignored.

Secondly, and more specifically, by uncovering the crucial role of selective *masking* in the case of the Bengalese Finch, we provide some support for Deacon's hypothesis and in doing so bring a new mechanism to the table for discussions of the origins of human syntax (though see also the discussion in [Deacon, 2003].) It is quite possible that we should not be looking for selective advantages of a culturally transmitted syntactic language, but rather asking what selective forces may have been shielded in our recent evolutionary past. The lifting of selection pressure, and the subsequent diversification of behaviour could have been the necessary precursors of a system of iterated learning for language. What remains to be understood is exactly what more is required for any subsequent modification and synergistic reorganisation of the neural mechanisms underlying these new behaviours.

We feel that computational modelling of the vocal behaviour of birds, humans and perhaps of other vocal learning species, may provide valuable insights to this question.

Acknowledgements

We would like to thank Kazuo Okanoya and Terry Deacon for very useful discussion on this work. We would also like to thank the members of the Language Evolution and Computation Research Unit at the University of Edinburgh for helpful comments and feedback. The first author is supported by an EPSRC studentship.

References

[Brighton, 2002] Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54.

[Bullock, 1999] Bullock, S. (1999). Are artificial mutation biases unnatural? In Floreano, D., Nicoud, J.-D., and Mondada, F., editors, *Fifth European Conference on Artificial Life (ECAL99)*, pages 64–73. Springer-Verlag.

[Catchpole and Slater, 1995] Catchpole, C. K. and Slater, P. J. B. (1995). *Bird Song: Biological themes and variations*. Cambridge University Press.

[Darwin, 1879] Darwin, C. (1879). *The Descent of Man, and Selection in Relation to Sex*. John Murray, London, 2nd edition. Reprinted in 2004 by Penguin.

[Deacon, 2003] 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, Cambridge, MA.

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

[Jarvis, 2004] Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, 1016:749–777.

[Kirby, 2001] 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):102–110.

[Kirby and Hurford, 2002] Kirby, S. and Hurford, J. R. (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 Verlag, London.

[Lachlan and Feldman, 2003] Lachlan, R. F. and Feldman, M. W. (2003). Evolution of cultural communication systems: the coevolution of cultural signals and genes encoding learning preferences. *Journal of Evolutionary Biology*, 16:1084–1095.

[Lachlan and Slater, 1999] Lachlan, R. F. and Slater, P. J. B. (1999). The maintenance of vocal learning by gene-culture interaction: the cultural trap hypothesis. *Proceedings of the Royal Society of London. B*, 266:701–706.

[Livingstone, 2002] Livingstone, D. (2002). The evolution of dialect diversity. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*, chapter 5, pages 99–118. Springer Verlag, London.

[Okanoya, 2002] Okanoya, K. (2002). Sexual display as a syntactic vehicle: The evolution of syntax in birdsong and human language through sexual selection. In Wray, A., editor, *The Transition to Language*, chapter 3. Oxford University Press, Oxford.

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

[Sasahara and Ikegami, 2004] Sasahara, K. and Ikegami, T. (2004). Song grammars as complex sexual displays. In *Artificial Life 9*.

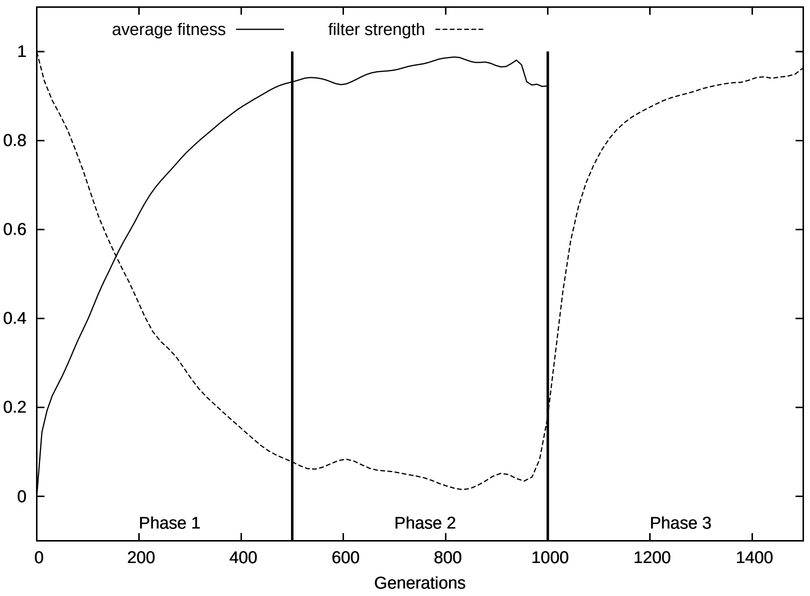
[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423 and 623–656. Reprinted in “Shannon Collected Papers”, ed. Sloane & Wyner, 1993, IEEE Press.

[Slater, 2003] Slater, P. J. B. (2003). Fifty years of bird song research: a case study in animal behaviour. *Animal Behaviour*, 65:633–639.

[Teal and Taylor, 2000] Teal, T. K. and Taylor, C. E. (2000). Effects of compression on language evolution. *Artificial Life*, 6:129–143.

[Wiles et al., 2005] Wiles, J., Watson, J., Tonkes, B., and Deacon, T. W. (2005). Transient phenomena in learning and evolution: Genetic assimilation and genetic redistribution. *Artificial Life*, (11):177–188.

[Zahavi, 1975] Zahavi, A. (1975). Mate selection – a selection for a handicap. *Journal of Theoretical Biology*, 53:205–214.



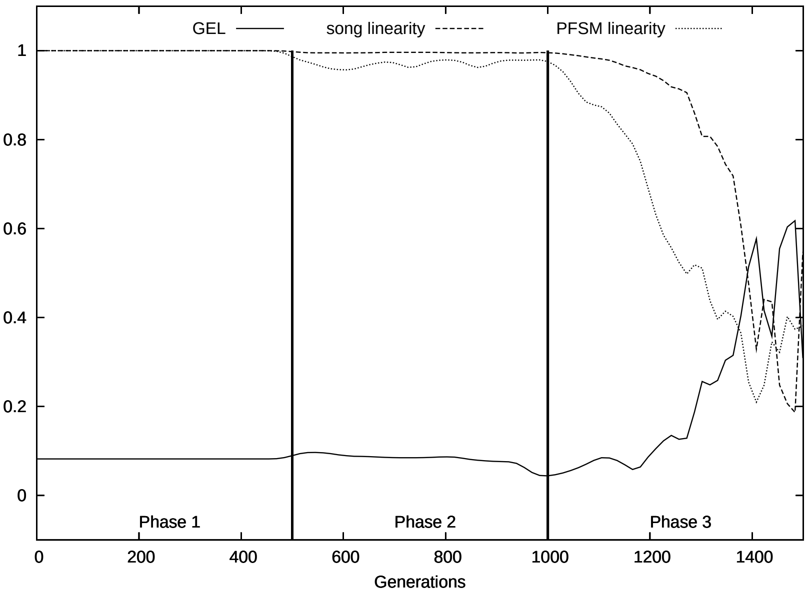


Fig. 1. These results are averages taken over 10 separate runs of the simulation with a different random number generator seed for each run. It should be noted that all these measures have been normalised into the range $0 - 1$. This graph therefore only shows the relative change each of the measures over the course of a run, not the absolute values of each measure. We have also smoothed the lines in the graph to better allow us to see the overall trends. More detailed results are available upon request. (PFSM is a probabilistic finite state machine, and GEL is the grammer encoding length of a PFSM.)

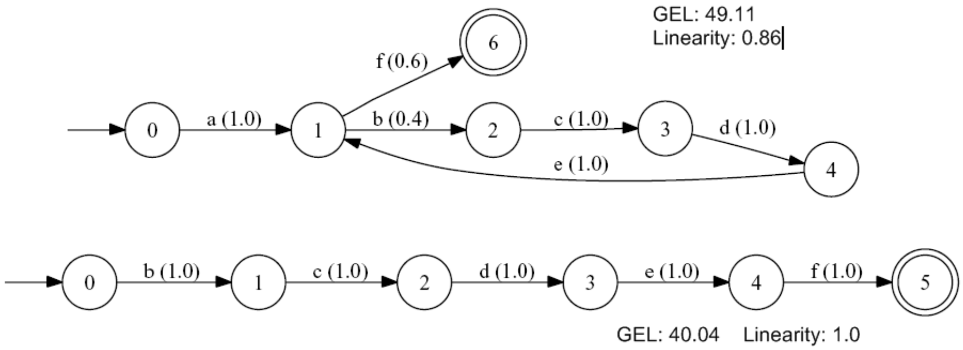


Fig. 2. Two example PFSMs from the population at the end of phase 2.

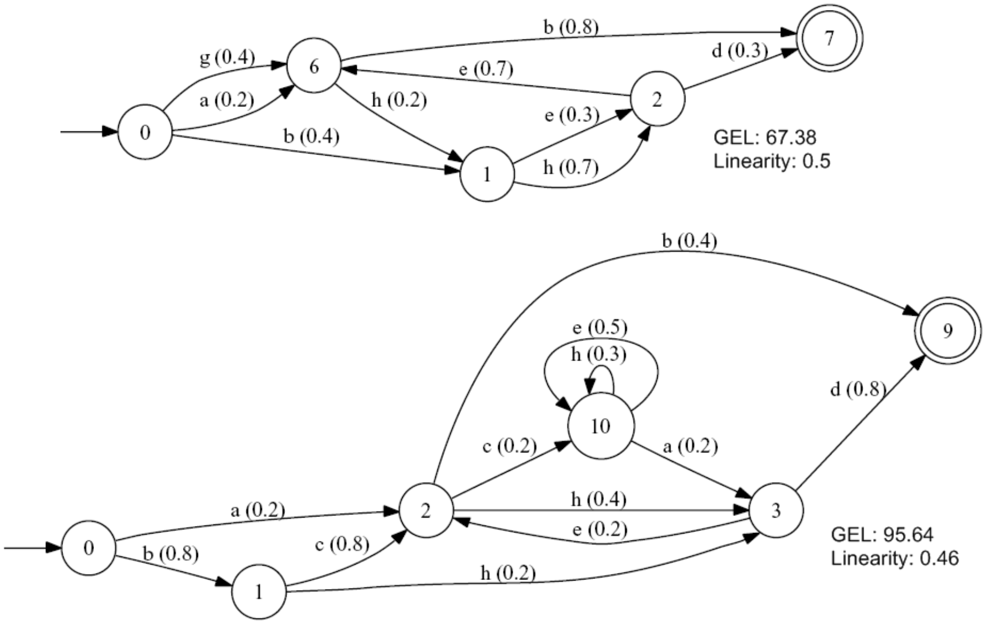


Fig. 3. Two example PFSMs from the population at the end of phase 3.

Available online at www.sciencedirect.com

ScienceDirect

Journal of Theoretical Biology 251 (2008) 570–583

Journal of Theoretical Biology

www.elsevier.com/locate/yjtbi

Song learning as an indicator mechanism: Modelling the developmental stress hypothesis

Graham R.S. Ritchie*, Simon Kirby, David J.C. Hawkey

Language Evolution and Computation Research Unit, University of Edinburgh, George Square, Edinburgh EH8 9LL, UK

Received 23 April 2007; received in revised form 29 November 2007; accepted 4 December 2007
Available online 27 December 2007

Abstract

The ‘developmental stress hypothesis’ attempts to provide a functional explanation of the evolutionary maintenance of song learning in songbirds. It argues that song learning can be viewed as an indicator mechanism that allows females to use learned features of song as a window on a male’s early development, a potentially stressful period that may have long-term phenotypic effects. In this paper we formally model this hypothesis for the first time, presenting a population genetic model that takes into account both the evolution of genetic learning preferences and cultural transmission of song. The models demonstrate that a preference for song types that reveal developmental stress can evolve in a population, and that cultural transmission of these song types can be stable, lending more support to the hypothesis.

Keywords: Developmental stress; Vocal learning; Bird song

1. Introduction

Bird song is a classic example of a sexually selected trait, and is also one of the best studied examples of non-human animal cultural transmission (reviewed in Catchpole and Slater, 1995; Marler and Slabbekoorn, 2004). Despite the wealth of studies and data on song function and development the evolutionary function of song *learning* remains unclear (Slater, 2003). There have been several proposed functions, such as the acquisition of a large and varied repertoire, habitat matching and neighbour matching (see, e.g. Marler and Slabbekoorn, 2004, for a review), but it seems that none of these apply widely to more than a few song learning species.

On the other hand, song learning appears to incur some costs, which suggests that the trait must have been positively maintained by selection (Lachlan and Slater, 1999). These costs include the development and maintenance of the complex system of brain nuclei which underlie song learning, and the substantial investment in time required to develop normal conspecific song. Learning also introduces the risk that a bird will fail to acquire a normal song due to a lack of suitable models; birds may acquire heterospecific song, or simply fail to find a tutor at all. Both of these will result in adult song that is unlikely to function as well as normal song. The ‘cultural trap hypothesis’, which argues that species that have developed cultural transmission of song are unlikely to lose the trait because of the interaction of genetic predispositions and processes of cultural evolution, is the only current hypothesis that can apply to a wide range of species (Lachlan and Slater, 1999; Lachlan and Feldman, 2003).

Recently, a more adaptive evolutionary function for song learning has been suggested which is in accord with several empirical studies. The ‘developmental stress hypothesis’ (DSH) proposes that song learning may serve as an indicator mechanism that allows females to judge a potential mate’s quality according to how well he has learned conspecific song (Nowicki et al., 1998; Buchanan et al., 2003; Nowicki and Searcy, 2004). Songbirds are altricial and infancy is fraught with potential stresses, including parasites and undernutrition (see, e.g. O’Connor, 1984). If developmental stress of a male bird has long-term effects, such as affecting the number of viable offspring

---

*Corresponding author. Tel.: +44 7905 258153.
E-mail address: graham.ritchie@ed.ac.uk (G.R.S. Ritchie).

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.12.013

a pairing can produce, it is plausible that selection would favour any mechanism that would allow females to accurately judge the degree of developmental stress of a potential mate. Various studies have shown that the brain structures mediating song learning develop in infancy, late compared to the rest of the avian nervous system, and so adult song behaviour may be irreparably affected by stresses that affect this development (discussed in Nowicki and Searcy, 2004). It seems plausible that females could evolve a preference for features of song that are acquired during this period as this would enable them to identify males affected by developmental stress.

The DSH has been supported by a range of empirical studies. A number of possible stressors have been studied in several species, for example; undernutrition has an effect on accuracy of song note copying in male swamp sparrows (*Melospiza georgiana*) (Nowicki et al., 2002a), social rank has an effect on acquired repertoire size in the European starling (*Sturnus vulgaris*) (Spencer et al., 2004) and food restriction and corticosterone administration selectively affect growth of the HVC (an important song control nucleus) in zebra finches (*Taeniopygia guttata*) (Buchanan et al., 2004). Nowicki et al. (2000) also provide support for a key prediction of the hypothesis, demonstrating that at least one measure of nestling growth correlates with adult repertoire size in the great reed warbler (*Acrocephalus arundinaceus*). Another prediction of the hypothesis is that females attend to these effects, and studies have shown that female song sparrows (*Melospiza melodia*) respond more to males that have copied the notes of their tutors more accurately (Nowicki et al., 2002b), and that female zebra finches significantly prefer the songs of control males to those of males stressed by either food restriction or corticosterone administration (Spencer et al., 2005).

These data suggest that several features of male song are affected by developmental stress and that females attend to these features, supporting some of the key predictions of the hypothesis. However, a number of issues raised remain unresolved; these include the effect of developmental stress on female preference and the cultural stability of a song type that is not acquirable by all members of the population. In this paper we attempt to address some of these issues and to add further support to this hypothesis using a population genetic model of the evolution of song preferences. With the model we investigate two important questions raised by the hypothesis: firstly, can a preference for culturally transmitted features of song affected by developmental stress evolve in a population, even when stress affects both female preference and male production? Secondly, can cultural transmission of songs with such features be stable, despite the fact that those members of the population suffering developmental stress cannot acquire or produce them?

2. The models

The models presented here are population genetic models in which we assume an infinite population with discrete generations and a 1:1 sex ratio. We also assume haploid genetics to simplify the analysis, as has been used in related models (e.g. Kirkpatrick, 1982; Aoki, 1989; Lachlan and Feldman, 2003).

We model developmental stress simply as resulting in either a high or low quality bird, whereby a low quality bird may be thought of as undernourished or parasitised. The probability of an individual bird of either sex being high quality is defined as δ. The factors affecting a bird's quality are assumed to be determined by the environment. Quality is therefore not inherited and is assigned to individuals independently of their genetic constitution and the quality of their parents. This is similar to Heywood's (1989) 'good' and 'bad' environments in his model of the handicap mechanism. We assume that it would be beneficial to females to be able to identify a potential mate's quality, but that a bird's quality is not directly observable. Instead we assume that the quality of a bird affects its ability to learn some song types. In the model, 'type 2' songs are only learnable by high quality birds, while 'type 1' songs are learnable by both high and low quality birds. Although we make no further assumptions about the song types other than this difference in learnability, 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. Considered in this way, we are therefore assuming that high quality birds are able to learn more complex songs than low quality birds. We also assume that only high quality birds are able to distinguish between the two song types. The proportion of type 2 songs in the cultural environment is denoted as t_2 and the proportion of type 1 songs $t_1 = 1 - t_2$. Importantly, we assume that both males and females are affected by developmental stress and that this stress can affect both sexes' adult behaviour.

A bird's ability to acquire a particular song type is also affected by a genetic component, we model this as a single autosomal locus A for which two alleles exist, A_1 and A_2.[1] This gene may be considered as coding for a learning bias and we assume (as did Lachlan and Feldman, 2003) that the same gene controls song learning bias in both males and females. This assumption seems plausible as studies have shown (reviewed in Riebel, 2003) that, in many species, both sexes are born with a preference for conspecific song which is affected by early song exposure (e.g. in zebra finches, Lauay et al., 2004), that females share many of the song system nuclei with males, and that females can produce song under abnormal hormonal conditions (Riebel, 2003). Birds with allele A_1 can only learn type 1 songs. Birds with allele A_2 are able to learn

---

[1] Although we only present results for a single locus here we have also modelled the case when song production and preference are controlled by genes at two separate loci. This makes algebraic analysis considerably harder (as we then have 16 different types of birds and also need to track the linkage disequilibrium between the two loci), but numerical iterations of the recursion equations indicate that the model's behaviour is qualitatively similar to the one locus case.

Table 1
The effect of the three factors on the adult phenotype

Quality (δ)	Allele (a_2)	Model song (P_{T_2})	Adult song	Adult preference	Label (X_i)	Frequency (x_i)
L	A_1	1	1	Either	X_1	$(1-\delta)(1-a_2)(1-P_{T_2}) = x_1$
H	A_1	1	1	Either	X_2	$\delta(1-a_2)(1-P_{T_2}) = x_2$
L	A_2	1	1	Either	X_3	$(1-\delta)a_2(1-P_{T_2}) = x_3$
H	A_2	1	1	Either	X_4	$\delta a_2(1-P_{T_2}) = x_4$
L	A_1	2	1	Either	X_5	$(1-\delta)(1-a_2)P_{T_2} = x_5$
H	A_1	2	1	Either	X_6	$\delta(1-a_2)P_{T_2} = x_6$
L	A_2	2	1	Either	X_7	$(1-\delta)a_2P_{T_2} = x_7$
H	A_2	2	2	2 (if $s>0$)	X_8	$\delta a_2P_{T_2} = x_8$

song types 1 and 2, but are biased to acquiring type 2 song. This bias will, however, only be expressed if a bird hears type 2 song as an infant. The proportion of the population with allele A_2 is denoted as a_2, with the proportion of birds with allele A_1, $a_1 = 1 - a_2$.

This gene affects adult song behaviour in both sexes in the following way. A high quality female with allele A_2 has a bias to acquiring type 2 song, and if she is exposed to this song type as an infant, she will memorise it and later will preferentially mate with a male singing a type 2 song. The strength of this mating preference is represented by the parameter s, with $s = 1$ implying such a female will exclusively mate with males singing type 2 song, and $s = 0$ meaning such a female mates indiscriminately, with intermediate values implying intermediate preference for type 2 song. All other females are unable to differentiate type 1 song from type 2 song and so will mate indiscriminately with a male singing either type. A high quality male with allele A_2 that is exposed to type 2 song is assumed to memorise this song and then go on to learn to produce it as an adult, i.e. we are assuming that the males follow the auditory template model of song development (see Catchpole and Slater, 1995, for a discussion of this model). Other males are only able to develop a type 1 song.

To model song acquisition, we assume that each infant bird is exposed to a set of m songs, randomly sampled from all the songs of the previous generation (this is thus a model of oblique cultural transmission as used in Lachlan and Feldman, 2003). They then pick a single model song to learn from. If the set contains at least one instance of type 2 song a high quality bird with allele A_2 will learn type 2 song. All other birds are unable to learn type 2 song, and will acquire type 1. The probability, P_{T_2}, of any bird hearing at least one instance of type 2 song is given by the following equation:

$$P_{T_2} = 1 - (1 - t_2)^m. \tag{2.1}$$

The following three factors thus contribute to an adult bird's phenotype:

- The bird's environmentally specified quality, either *high* (denoted as H) or *low* (denoted as L).
- The allele it inherits from its parents, either A_1 or A_2.
- The set of songs the bird hears as an infant.

The effects of these factors on the adult song and song preference behaviour is shown in Table 1. There are eight possible combinations of quality, allele and model song, giving eight types of individual of each sex in the model which we label $X_1, \ldots, X_8$. The frequencies of these types are given by the values of δ, a_2 and t_2 in each generation, as shown in the table, which we denote as x_i. In summary, only high quality individuals bearing allele A_2 who are exposed to type 2 song, identified as type X_8 in the table, will develop a mating preference for type 2 song (as females) and are able to produce type 2 song (as males). All other individuals show no adult mating preference and can only produce type 1 song. Type 2 song is thus a reliable indicator of male quality; only high quality males are able to acquire and later produce it.

With this model we are interested in the conditions under which allele A_2, which governs the ability to acquire type 2 song, can spread in a population even when it has no effect on some carriers of the allele (namely types X_3, X_4 and X_7), and reduces the number of potential mates for the choosy X_8 females (assuming that $s > 0$). We are also interested in whether cultural transmission of the type 2 song can be stable despite the fact that it cannot be acquired or produced by every individual in the population.

2.1. Polygyny

For the first full model, we assume a polygynous mating system. We are aware that many songbird species are largely monogamous and we model this in Section 2.2, but as these models can be seen as mathematical variations on the polygynous model, for clarity we first present the polygynous case. This also allows comparison with other models of the evolution of culturally transmitted song which also assume polygyny (e.g. Aoki, 1989; Lachlan and Feldman, 2003).

We assume that each time a female chooses a mate she samples n males and will mate with the male whose song she prefers the most.[2] The X_8 females prefer to mate with

---

[2] This is thus modelling a 'best of n' mate selection strategy, which is one of a number of possible mate selection strategies. While we have not formally incorporated any other strategies into the model, we feel that as long as the key property that the choosy X_8 females are not guaranteed to find a preferred mate remains, the results would be qualitatively similar.

Table 2
Frequencies of matings for the polygynous model

Females	Males			
	X_1	$\dots$	X_7	X_8
X_1				$x_1 \cdot x_8$
$\vdots$		$x_i \cdot x_j$		$\vdots$
X_7				$x_7 \cdot x_8$
X_8	$(1-s)\cdot x_8 \cdot x_1$	$\dots$	$(1-s)\cdot x_8 \cdot x_7$	$((1-s)\cdot x_8 \cdot x_8) + (s \cdot x_8 \cdot P_{X_8})$

X_8 males, and a proportion s of the time such a female will exclusively pick an X_8 male from their set of n, if there is not an X_8 male in this set they will not mate. The probability, P_{X_8}, that at least one X_8 male is in the sample of n males is given by the following equation:

$$P_{X_8} = 1-(1-x_8)^n. \tag{2.2}$$

A proportion $1-s$ of the time the X_8 females, as well as all other females, mate indiscriminately, but will always find a mate. The frequencies of matings between the different types of birds under these assumptions is given in Table 2. As we are assuming polygyny, each male is able to mate multiple times.

In the model, we assume that the effect of quality on fitness is to affect the number of viable offspring a mated pair produce. This seems like a plausible assumption, as a parasitised or undernourished male is likely to be a less effective parent than a high quality male. Unfortunately we are not aware of any direct empirical investigation of the effect of developmental stress on male parental abilities, but Naguib et al. (2006) show that the developmental stress (in their study, brood size) of mother zebra finches has an effect on the hatching and fledgling success of their daughters one generation later. We do, however, have some indirect evidence for the effect on male parental abilities. Hasselquist (1998) shows that male syllable repertoire size is positively correlated with harem size and fledging success in great reed warblers, and, as mentioned earlier, Nowicki et al. (2000) show that at least one measure of nestling developmental stress has an effect on syllable repertoire size in this species, so it seems that developmentally stressed males may fare worse as parents. In reality, developmental stress is also likely to affect an individual's chance of survival until reproductive age. However, with this model we are only concerned with the effect of quality on reproductive success, and so we assume that birds of high and low quality are equally viable until reproductive age.

We define the fitness of mated pairs as a function of the qualities of both the female and her chosen mating partner. The fitness of a bird that does not mate is assumed to be 0. There are four classes of possible matings, a high or low quality female mating with a high or low quality male. The relative fitnesses of each of these classes of pairings are defined in Table 3. This value can be considered a measure of the relative number of viable offspring that a mated pair will produce, and we assume here that these values will take a value between 0 and 1. Alleles A_1 and A_2 are equally likely to be transmitted from an $A_1 \times A_2$ mating.

Table 3
The relative fitness of each of the four possible classes of mated pairs as a function of individual quality

Female	Male	
	H	L
H	1	f_{HL}
L	f_{LH}	f_{LL}

According to these assumptions we can now write recursion equations for the value of a_2 in the next generation, a_2', and the value of t_2 in the next generation, t_2', which is simply the proportion of X_8 males in the current generation. These recursions therefore include both genetic and cultural transmission. The equations demonstrate that for allele A_2 to spread, the average fitness of birds with this allele (w_{A_2}) must be greater than the average fitness of birds with allele A_1 (w_{A_1}):

$$t_2' = x_8, \tag{2.3}$$

$$a_2' = \frac{w_{A_2}}{w_{A_1} + w_{A_2}}, \tag{2.4}$$

where

$$\begin{aligned} w_{A_2} = {} & f_{LL} \cdot [(x_3+x_7)(x_1+x_3+x_5+x_7)] \\ & + f_{LH} \cdot \Bigg[(x_3+x_7)(x_4+x_8) \\ & + \frac{(x_1+x_5)(x_4+x_8)+(x_3+x_7)(x_2+x_6)}{2}\Bigg] \\ & + f_{HL} \cdot \Bigg[((1-s)\cdot x_8 + x_4)(x_3+x_7) \\ & + \frac{(x_2+x_6)(x_3+x_7)+((1-s)\cdot x_8+x_4)(x_1+x_5)}{2}\Bigg] \\ & + 1 \cdot \Bigg[s \cdot x_8 \cdot P_{X_8} + (1-s)\cdot x_8 \cdot (x_4+x_8) \\ & + x_4 \cdot (x_2+x_4+x_6+x_8) \\ & + \frac{(1-s)\cdot x_8 \cdot (x_2+x_6)+(x_2+x_6)\cdot x_8}{2}\Bigg], \end{aligned}$$

$$\begin{aligned} w_{A_1} &= f_{LL}\cdot[(x_1+x_5)(x_1+x_3+x_5+x_7)] \\ &\quad + f_{LH}\cdot\left[(x_1+x_5)(x_2+x_6) + \frac{(x_1+x_5)(x_4+x_8)+(x_3+x_7)(x_2+x_6)}{2}\right] \\ &\quad + f_{HL}\cdot\left[(x_2+x_6)(x_1+x_5) + \frac{(x_2+x_6)(x_3+x_7)+((1-s)\cdot x_8+x_4)(x_1+x_5)}{2}\right] \\ &\quad + 1\cdot\left[(x_2+x_6)(x_2+x_4+x_6) + \frac{(1-s)\cdot x_8\cdot(x_2+x_6)+(x_2+x_6)\cdot x_8}{2}\right]. \end{aligned}$$

We can graphically represent the co-evolution of the A_2 allele and type 2 song with a vector-field plot in two dimensions which shows how a_2 and t_2 change from a range of possible values, this is shown in Fig. 1. This figure demonstrates that for all values of a_2 when $t_2 \neq 0$ there is frequency-dependent selection for or against allele A_2, except for an internal equilibrium. While the precise value of a_2 above which allele A_2 invades depends on the value of t_2, the vector plots show that it is approximately true that when a_2 is above the value at the internal equilibrium, A_2 invades. When a_2 is below this value, generally A_2 is selected against. Thus we use the value of a_2 at the internal equilibrium as an indicator of the value of a_2 above which A_2 will invade.

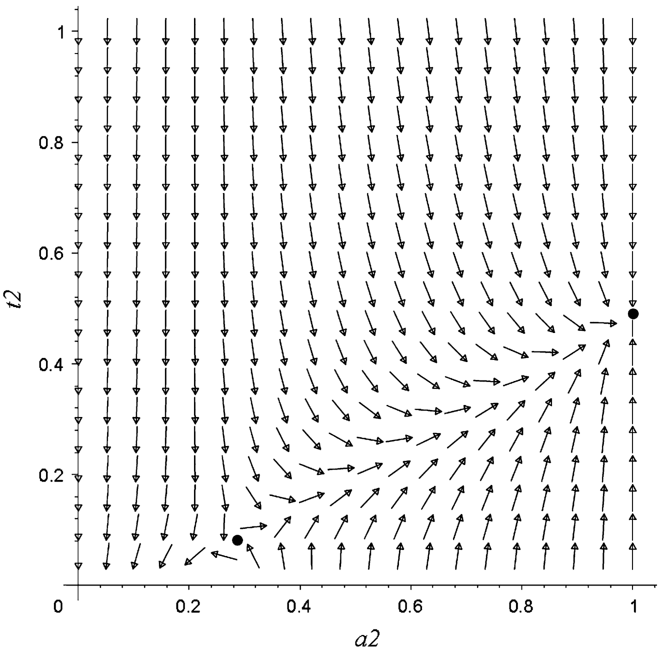


Fig. 1. Vector-field plot showing how the frequency of the type 2 song preference allele, a_2, and the frequency of type 2 song in the cultural environment, t_2, change from a range of possible values in the polygynous model. The equilibrium points are marked with a circle. The parameter values used here are: $\delta = 0.5$, $m = 10$, $n = 5$, $f_{LH} = 0.75$, $f_{HL} = 0.5$, $f_{LL} = 0.25$, $s = 0.75$. Note that the maximum possible value of t_2 is given by δ so at the $\hat{a}_2 = 1$ equilibrium in this graph $t_2 = 0.5$.

At equilibrium $a_2' = a_2$ and $t_2' = t_2$. The recursion equations (2.3) and (2.4) can be solved to give the values at which these conditions hold. The equilibrium equations do not have a general analytic solution (because of terms raised to the powers of the parameters m and n), but we can look at extremes of the model to get some understanding of the general behaviour. If we assume that $t_2 > 0$ and that m is large then the probability of hearing a type 2 song, $P_{T_2} \approx 1$. Substituting this into the equations means we find two equilibria where A_2 is either at fixation in the population or is extinct. The equilibrium values for a_2 and t_2 at these points are; $\{\hat{a}_2 = 0, \hat{t}_2 = 0\}$, $\{\hat{a}_2 = 1, \hat{t}_2 = \delta\}$. As the vector-field plot suggests, there is no stable equilibrium where alleles A_1 and A_2 coexist and $t_2 \neq 0$. The expression for the unstable polymorphic equilibrium is rather complicated and so is included in Appendices A–C.

The equilibrium we have found when $\hat{a}_2 = 0$ is not actually a valid equilibrium of the full model, as when $a_2 = 0$, $t_2 = 0$, so our assumption that $P_{T_2} \approx 1$ no longer holds. In fact solving the full model finds one solution when $t_2 = 0$, $\{\hat{a}_2 = a_2, \hat{t}_2 = 0\}$. This means that when $t_2 = 0$ the value of a_2 will not change. This is an intuitive result as when there are no type 2 songs in the environment the birds carrying A_2 behave in exactly the same way as the A_1 carrying birds, and so selection cannot act on the frequency of A_2.

A full stability analysis of these three equilibria is included in Appendices A–C, and this shows that the A_2 fixation equilibrium is the only generally stable equilibrium when $s > 0$. The polymorphic equilibrium is unstable for a wide range of the parameter values, and the equilibrium where $t_2 = 0$ is stable to perturbations in t_2 for some values of a_2 and the parameters, but perturbations in a_2 will persist. This has some interesting consequences for a finite population.

When $t_2 = 0$ selection cannot act on a_2. This means that in a finite population the value of a_2 will be controlled by effects such as pleiotropy and drift. If a_2 were to drift to a value greater than the polymorphic equilibrium value and there was then some cultural invention or mutation event which created some type 2 songs, allele A_2 could then invade the population and remain stable.[3] As the stability analysis shows, invasion of allele A_2 is the only globally stable equilibrium value, but the likelihood of invasion in a finite population will increase as the polymorphic equilibrium value of a_2 decreases, i.e. as the basin of attraction for the invasion of A_2 gets larger.

The polymorphic equilibrium value of a_2 depends on only four of the model parameters (m, n, δ and f_{HL}, for

---

[3] As noted earlier, this is not a completely rigorous analysis as there are some values of a_2 larger than the equilibrium value that will be selected against for low values of t_2, but the value of a_2 at the polymorphic equilibrium gives us a reasonable numerical proxy for values above which A_2 will be positively selected for.

details of this expression see Appendices A–C). To demonstrate the effects of these parameters on the polymorphic equilibrium value of a_2, and hence the opportunity for invasion of A_2, we varied one parameter at a time while holding the others constant. The effects of each of these parameters are shown in Fig. 2. These graphs demonstrate that as the number of songs heard in infancy (m), the number of males each female chooses a mate from (n), and the proportion of high quality individuals (δ) increase, the value of a_2 above which A_2 is selected for decreases. As the relative fitness of high quality females who mate with low quality males (f_{HL}) increases it becomes harder for allele A_2 to invade. These results bear out the intuition that as factors increasing the chance that an X_8 female finds an X_8 mate increases, invasion by A_2 is easier. When the relative fitness of high quality females who mate with low quality males increases, the advantage of the X_8 females over other high quality females decreases and so the opportunity for invasion of A_2 decreases.

Some of the model parameters do not affect the equilibrium values but do have an effect on the stability of the equilibria. These parameters thus affect the likelihood of finding the system at a given equilibrium and the speed at which the system evolves to an equilibrium. As Appendices A–C demonstrate in detail, increasing the relative fitness of a low quality female mating with a low quality male (f_{LL}) and that of a low quality female mating with a high quality male (f_{LH}) reduces the advantage of X_8 females and so decreases the stability of invasion by A_2. Increasing the strength of the X_8 female's preference for type 2 song singing males (s) increases the disparity in behaviour between the X_8 females and the other types and so will speed up the invasion or extinction of allele A_2. When $s = 0$ there are no stable equilibria and there is no selection for or against A_2.

2.2. Monogamy

Many songbird species are largely monogamous and, as our model assumes that males only provide direct benefits to females (e.g. food, parental care), our assumption in the previous section that high quality males may mate an unlimited number of times may be misleading. It is unlikely that a high quality male would be able to provide such benefits to several females at once. In this section we therefore incorporate monogamy into the model.

To model monogamy we need to assume that as a female chooses a mate this male is no longer available for mating with any other female. When $s > 0$, a proportion s of the X_8 females (henceforth referred to as the 'choosy' X_8 females) preferentially choose X_8 males and when they make their choices they will skew the distribution of available X_8 males in the population. All other females choose at random and so they will not change the distribution of males. We have been unable to deal with this mathematically in general and instead we have made two simplifying assumptions to allow us to calculate the mating frequencies. Firstly we follow O'Donald (1980) in that we only consider the two cases when either the choosy proportion of the X_8 females get to pick their mates *before* all the other females, or when these females choose *after* all the other

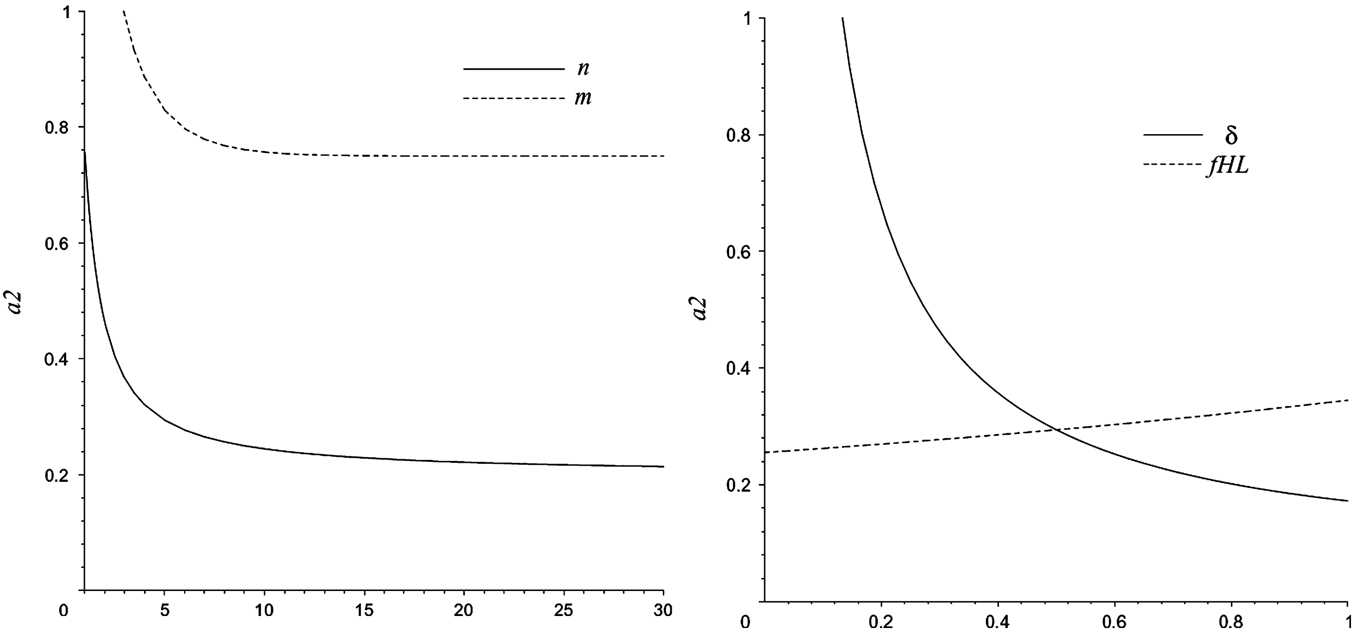


Fig. 2. Graphs showing the effect of the number of songs heard as an infant (m), the number of males a female chooses a mate from (n), the proportion of high quality birds (δ), and the relative fitness of a high quality female mating with a low quality male (f_{HL}) on the polymorphic equilibrium value of a_2 in the polygynous model. The default parameter values used are: $\delta = 0.5$, $m = 10$, $n = 5$, $f_{HL} = 0.5$. Note that increasing m, n and δ decreases the equilibrium value, making invasion by A_2 more likely, while increasing f_{HL} makes invasion less likely.

Table 4
Frequencies of matings for the monogamous model when the choosy $s \cdot x_8$ proportion of the X_8 females choose before the other females

Females	Males			
	X_1	...	X_7	X_8
X_1				$\frac{x_1 \cdot (1-s) \cdot x_8}{1-(s \cdot x_8)}$
$\vdots$		$\frac{x_i \cdot x_j}{1-(s \cdot x_8)}$		$\vdots$
X_7				$\frac{x_7 \cdot (1-s) \cdot x_8}{1-(s \cdot x_8)}$
X_8	$\frac{(1-s) \cdot x_8 \cdot x_1}{1-(s \cdot x_8)}$	...	$\frac{(1-s) \cdot x_8 \cdot x_7}{1-(s \cdot x_8)}$	$\frac{((1-s) \cdot x_8)^2}{1-(s \cdot x_8)} + (s \cdot x_8)$

Table 5
Frequencies of matings for the monogamous model when the choosy $s \cdot x_8$ proportion of the X_8 females choose after the other females

Females	Males			
	X_1	...	X_7	X_8
X_1				$x_1 \cdot x_8$
$\vdots$		$x_i \cdot x_j$		$\vdots$
X_7				$x_7 \cdot x_8$
X_8	$(1-s) \cdot x_8 \cdot x_1$	...	$(1-s) \cdot x_8 \cdot x_7$	$((1-s) \cdot x_8 \cdot x_8) + (s \cdot x_8 \cdot x_8)$

females.[4] The former approach is the one taken by some previous models of monogamy (e.g. Andersson, 1986), but we model both extremes as these represent the highest and lowest possible number of matings involving X_8 females. These extremes therefore give us an upper and lower bound on the case where the X_8 females choices are intermixed with the other females. The second simplification we make is that the number of mates females choose from, $n = \infty$ so that a choosy X_8 female is guaranteed to find an X_8 male if there are any left in the population (as again otherwise we would have to deal with a changing distribution of male types). We realise that both of these assumptions are biologically implausible, but we hope that these extremes of the model will provide some insight on the full behaviour that we could not otherwise obtain.

The frequencies of matings between all eight phenogenotypes when the X_8 females choose first is given in Table 4, and the frequencies when the X_8 females choose last is given in Table 5.

For the case where the choosy X_8 females choose after the other females, a proportion $(1-x_8)+(1-s) \cdot x_8$ of the X_8 males will have paired off with other females. This will leave a proportion $s \cdot x_8$ of the X_8 males, so the frequency of choosy $X_8 \times X_8$ matings will be $s \cdot x_8 \cdot x_8$ and the remaining proportion $1-x_8$ of the choosy X_8 females will not find a mate. This means that this model is in fact a special case of the polygynous model for $P_{X_8} = x_8$ (which is mathematically equivalent to setting $n = 1$) and so the recursion equations (2.3) and (2.4) and the equilibria found earlier are the same for this model, but with n set to 1.

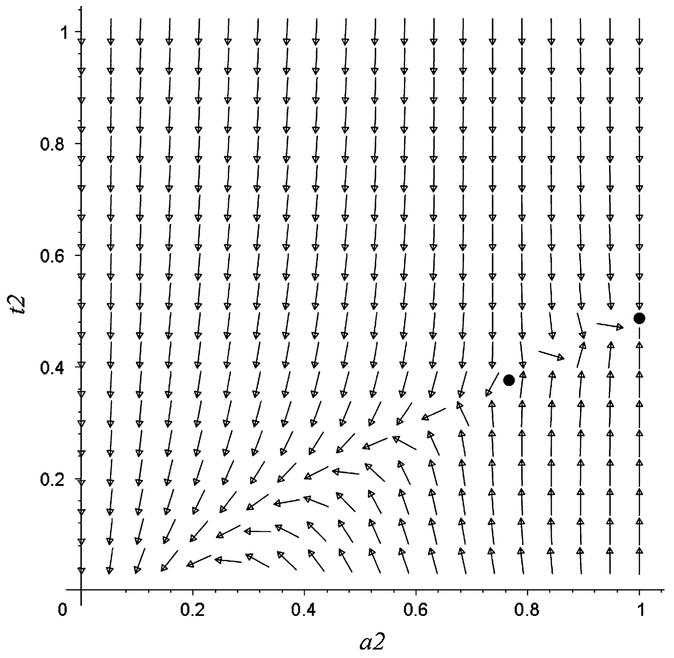


Fig. 3. Vector-field plot for the monogamous model when the X_8 females choose after the other females. The equilibrium points are marked with a circle. The parameter values used here are: $\delta = 0.5$, $m = 10$, $f_{LH} = 0.75$, $f_{HL} = 0.5$, $f_{LL} = 0.25$.

---

[4] We do not consider the proportion $(1-s)$ of the X_8 females as they choose indiscriminately and therefore do not affect the distribution of males.

A vector-field plot for this model is shown in Fig. 3. We see that the internal equilibrium for this model (with the same values of m, δ and f_{HL} as used in the polygynous model) requires a higher value of a_2 and so we would expect that allele A_2 would be less likely to invade this population than for the polygynous model. Fig. 5 shows the effect of m, δ and f_{HL} on the polymorphic equilibrium value of a_2, again demonstrating that higher values of m and δ increase the chance of invasion of A_2. Increasing f_{HL} again decreases the chance of invasion by A_2 but it has a rather stronger effect here than in the polygynous model. We can see that when $f_{HL} = 1$ there is no region of selection for A_2 and so it would be very unlikely to invade. In other words, as expected, there needs to be some advantage for high quality females to mate with high quality males.

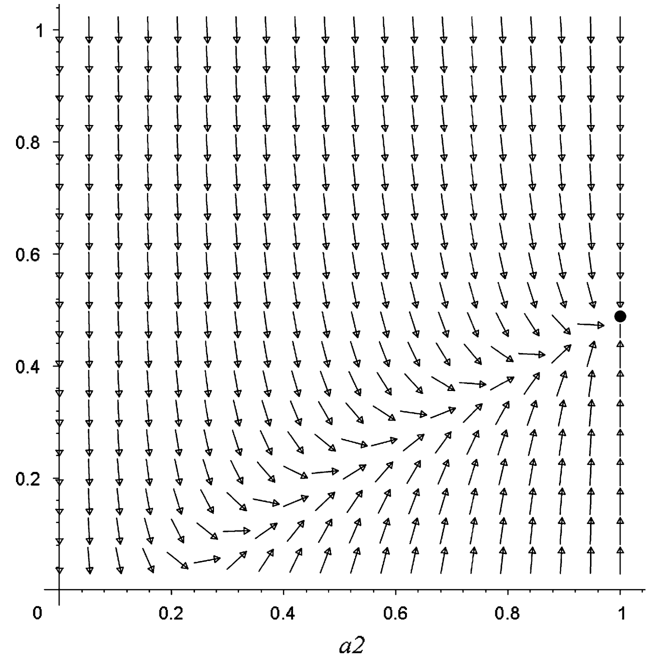


Fig. 4. Vector-field plot for the monogamous model when the X_8 females choose before the other females. The equilibrium point is marked with a circle. Note that there is no polymorphic equilibrium in this model. The parameter values used here are: $\delta = 0.5$, $m = 10$, $f_{LH} = 0.75$, $f_{HL} = 0.5$, $f_{LL} = 0.25$.

For the case where the choosy X_8 females choose before the other females we note that, by definition, there will always be as many X_8 males as X_8 females. As we assume here that $n = \infty$, this means that the choosy proportion s of the X_8 females are guaranteed to be paired with an X_8 male so there remains only a proportion $1 - s$ of the X_8 males remaining for the other females to select. We therefore need to redefine the recursion equations slightly for this model. Essentially we reduce the frequency of any mating event that includes an X_8 male mating with any female type other than a choosy X_8 female by $1 - s$, and we then divide each of the remaining mating event frequencies by $1 - (s \cdot x_8)$ to allow for the reduced frequency of X_8 males. The full recursions for this model are given in Appendices A–C and a vector-field plot for this model is shown in Fig. 4. This model also has a stable equilibrium with A_2 at fixation, and again, if $t_2 = 0$ any value of a_2 is stable. Unlike the other two models there is no polymorphic equilibrium and there is no region of selection against A_2, this is a consequence of the guarantee that a choosy X_8 female will find a high quality X_8 mate; i.e. unlike the previous models there is no risk for being choosy (Fig. 5).

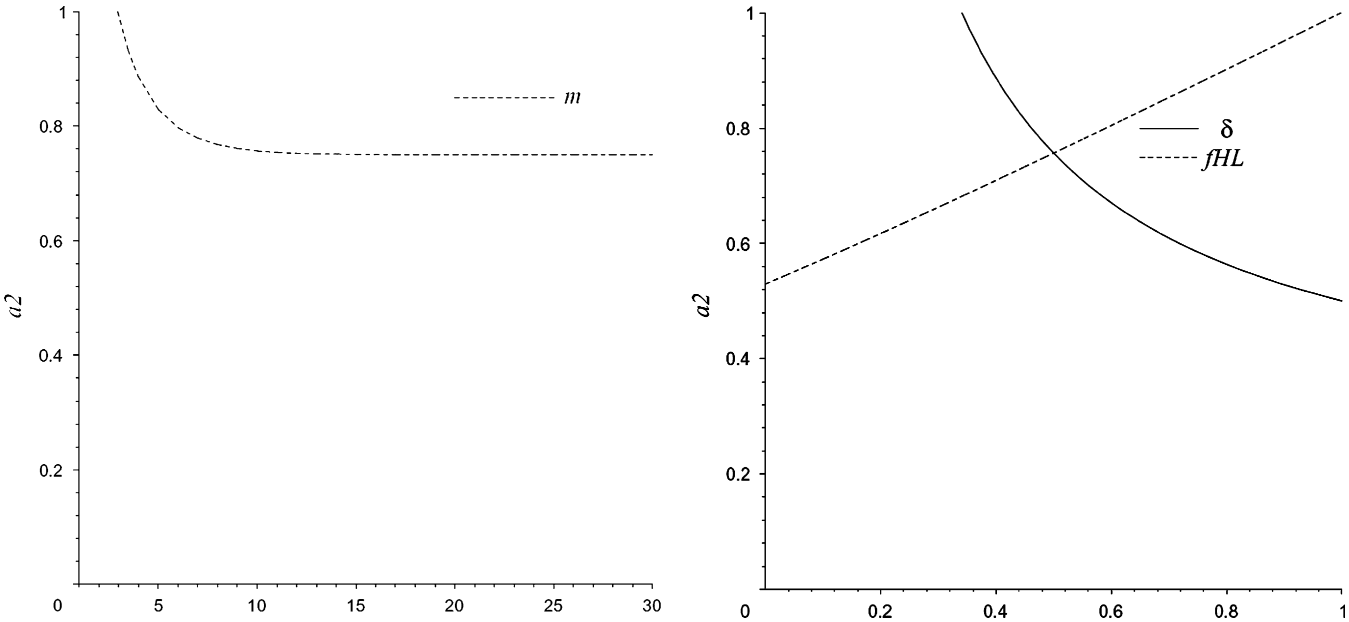


Fig. 5. Graphs showing the effect of the number of songs heard as an infant (m), the proportion of high quality birds (δ), and the relative fitness of a high quality female mating with a low quality male (f_{HL}) on the polymorphic equilibrium value of a_2 in the monogamous model when the X_8 females choose after the other females. The default parameter values used are: $\delta = 0.5$, $m = 10$, $f_{HL} = 0.5$. Note that increasing m and δ decreases the equilibrium value, making invasion by A_2 more likely, while increasing f_{HL} makes invasion less likely.

Although there is no region of selection against A_2 in this model there is a region where there is no selection for A_2. As the stability analysis included in Appendices A–C demonstrate, A_2 will be subject to selection when $a_2 > 1/\delta m$. When this condition is satisfied a small perturbation to t_2 will allow A_2 to be positively selected for. Overall in this model then, we would expect allele A_2 to be able to invade this population more easily than in the previous two models but the likelihood of invasion is again dependent on the values of δ and m, with larger values making invasion more likely.

As demonstrated in Appendices A–C, for both of these models, the parameters f_{LL}, f_{LH} and s have the same qualitative effect on the stability of the equilibria as we found for the monogamous model.

These models thus provide the first theoretical support for the DSH, showing that a preference for less learnable 'complex' song can evolve, and that cultural transmission of this song type is stable.

3. Discussion

The DSH proposes that learned features of male song can be used as an indicator of early developmental stress. The models we present demonstrate 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. They also show that there can be stable cultural transmission of these less learnable song types. The opportunity for invasion of such an allele in a finite population depends; however, on several biologically relevant parameters, including the prevalence of stressors, the number of unique songs heard as an infant and the number of potential mates available to females to choose from. Assumptions about the mating system also have an important effect. The models also incorporate some considerations that have yet to be investigated empirically, such as the effect of early song exposure and developmental stress on female preferences.

We make very few assumptions about the nature of the song types, so any feature of song whose acquisition is affected by developmental stress may be used as an indicator. This seems consistent with the biological data, because, as we discussed earlier, it appears that females of several different species attend to different features of song, but importantly all of which seem to be particularly affected by developmental stress.

The models presented here are admittedly simple, and there are a number of modifications we could make to make them more biologically plausible, such as working with a finite population and modelling the cultural transmission of song in a more realistic way than the panmictic cultural environment we assume here. Making such modifications would, however, make the mathematical analysis of the interactions of the parameters considerably harder and less clear. Nevertheless, we feel that these simple models incorporate some of the key features of the DSH and, while we hope to relax some of these assumptions in future work, as they stand, the models suggest a number of predictions which may be testable empirically and we present some here.

Firstly, if there are a low proportion of high quality birds, i.e. the average level of developmental stress in the population is high, the model predicts that we will be less likely to see such a system evolve. This is because it will be harder for the few high quality birds to overcome the frequency-dependent selection against them below the threshold value (which we discuss further below). The model also predicts that we are more likely to see such a system evolve in species where the females are able to pick mates from a relatively large pool of males, and, in monogamous species, where the high quality females are able to pick mates earlier in the breeding season than low quality females (as originally suggested by Darwin, 1879). A final, rather intuitive, prediction is that if the fitness detriment of mating with a low quality male is small then there is little advantage to finding a high quality male and such a system is much less likely to evolve.

Unlike other models of sexual selection, we do not include an explicit cost difference for birds singing or preferring either of the two song types. This was a deliberate decision, as it appears that many of the features of male song that female birds have been shown to prefer appear to be relatively cheap to produce (Nowicki and Searcy, 2004). As an example, it is not clear that a bird singing a song with a larger range of acquired syllables will, for example, expend more time or energy, or be more likely to be predated upon, than a bird singing a simpler song. Yet, as mentioned earlier, female great reed warblers appear to prefer males singing songs with more syllable variation, and males singing in this way appear to be more successful parents (Hasselquist, 1998). It is generally accepted that in order to be reliable a signal must be costly in some way (e.g. Grafen, 1990) and so this appears to be something of a quandary. The original intention of Nowicki et al. (1998) in proposing the DSH was to resolve this problem. They suggested that learned features of song can be reliable indicators of male quality as a result of the *developmental* cost of acquiring the song. Our models include this developmental cost implicitly by assuming that high quality individuals are able to learn some song types that low quality individuals cannot. In assuming that the X_8 females will sometimes fail to find a mate, we also impose a cost on choosiness.

In the first two models we found a strong threshold effect, due to the frequency-dependent selection for or against X_8 individuals. We found that the frequency of the A_2 allele must rise above a threshold value, by some mechanism such as random drift, before it can stably invade the population. How likely this situation is to arise is therefore significant in our analysis of the model. All we require of such an allele is that it makes the bearer predisposed to learning some song types rather than others; this might be as a result of some form of auditory bias that occurred by mutation or drift, or by selection on auditory capacities not related to song preference. We feel that such mutations are likely to have occurred fairly frequently in the evolution of the songbirds to give rise to the great variety of song types and preferences we find today. The only special property we require of this bias is that the ability to learn a song that satisfies the bias is correlated in some way with the level of developmental stress. As we discussed in Section 2.1, 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. High quality bearers of the allele are then able to reliably find a high quality mate and, in sufficient numbers, the allele can stably invade the population.

Previous theoretical work on the evolution of cultural communication systems, especially human language, (e.g. Brighton et al., 2005; Oudeyer, 2005) has argued that there is often a cultural evolutionary pressure in such systems for the signals to become increasingly learnable. This seems an intuitive result as culturally transmitted communication systems must be able to be acquired by each new generation of learners if they are to persist, and so less learnable forms are likely to be selected against. It is a surprising result of our models that selection can favour a preference for a song type that is, by definition, less learnable, and that such types can be stably culturally transmitted. This is a result of the fact that less learnable song types allow female birds to differentiate males according to their learning ability. The idea that songs that are less learnable might be more attractive to females is supported by recent work showing that both female swamp sparrows (Ballentine et al., 2004) and canaries (*Serinus canaria*) (Draganoiu et al., 2002) prefer songs that are closer to the physical performance limit of males (in these cases of the trill rate and frequency bandwidth), and therefore hardest to produce and, plausibly, to acquire.

The models synthesise the available biological data and verbal arguments and provide the first theoretical support for a novel role for cultural transmission in sexual selection (though see also Laland, 1994). This may have implications for other species that rely on cultural transmission, for example humans are also altricial and acquire language early in life. 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.

Acknowledgements

We thank Andy Gardner, Nick Barton, Richard Carter, Thom Scott-Philips, Stefan Hoefler, other members of the LEC at Edinburgh, and two anonymous referees for very useful comments on earlier drafts. The first author was supported by an EPSRC studentship from the School of Informatics at the University of Edinburgh.

Appendix A. The expression for the polymorphic equilibrium for the polygynous model

$$\hat{a}_2 = \frac{e^{(Y/n)} - 1}{\delta \cdot (e^{(Ym/n)} - 1)}, \tag{A.1}$$

$$\hat{t}_2 = 1 - e^{(Y/n)}, \tag{A.2}$$

where

$$Y = \ln\left(1 + \frac{f_{HL} \cdot \delta - f_{HL} - \delta}{2}\right).$$

Appendix B. Recursion equations for the monogamous model when the X_8 females choose before the other females

$$t_2' = x_8, \tag{B.1}$$

$$a_2' = \frac{w_{A_2}}{w_{A_1} + w_{A_2}}, \tag{B.2}$$

where

$$\begin{aligned} w_{A_2} = {} & f_{LL} \cdot \left[\frac{(x_3 + x_7)(x_1 + x_3 + x_5 + x_7)}{1 - (s \cdot x_8)}\right] \\ & + f_{LH} \cdot \left[\frac{(x_3 + x_7)(x_4 + (1 - s) \cdot x_8)}{1 - (s \cdot x_8)} + \frac{(x_1 + x_5)(x_4 + (1 - s) \cdot x_8) + (x_3 + x_7)(x_2 + x_6)}{2 \cdot (1 - (s \cdot x_8))}\right] \end{aligned}$$

$$
\begin{aligned}
&+f_{HL}\cdot\left[\frac{((1-s)\cdot x_8+x_4)(x_3+x_7)}{1-(s\cdot x_8)}+\frac{(x_2+x_6)(x_3+x_7)+((1-s)\cdot x_8+x_4)(x_1+x_5)}{2\cdot(1-(s\cdot x_8))}\right]\\
&+1\cdot\left[s\cdot x_8+\frac{(1-s)\cdot x_8\cdot(x_4+(1-s)\cdot x_8)+x_4\cdot(x_2+x_4+x_6+(1-s)\cdot x_8)}{1-(s\cdot x_8)}\right.\\
&\left.+\frac{(1-s)\cdot x_8\cdot(x_2+x_6)+(x_2+x_6)\cdot(1-s)\cdot x_8}{2\cdot(1-(s\cdot x_8))}\right],
\end{aligned}
$$

$$
\begin{aligned}
w_{A_1}=&f_{LL}\cdot\left[\frac{(x_1+x_5)(x_1+x_3+x_5+x_7)}{1-(s\cdot x_8)}\right]\\
&+f_{LH}\cdot\left[\frac{(x_1+x_5)(x_2+x_6)}{1-(s\cdot x_8)}+\frac{(x_1+x_5)(x_4+(1-s)\cdot x_8)+(x_3+x_7)(x_2+x_6)}{2\cdot(1-(s\cdot x_8))}\right]\\
&+f_{HL}\cdot\left[\frac{(x_2+x_6)(x_1+x_5)}{1-(s\cdot x_8)}+\frac{(x_2+x_6)(x_3+x_7)+((1-s)\cdot x_8+x_4)\cdot(x_1+x_5)}{2\cdot(1-(s\cdot x_8))}\right]\\
&+1\cdot\left[\frac{(x_2+x_6)(x_2+x_4+x_6)}{1-(s\cdot x_8)}+\frac{(1-s)\cdot x_8\cdot(x_2+x_6)+(x_2+x_6)\cdot(1-s)\cdot x_8}{2\cdot(1-(s\cdot x_8))}\right].
\end{aligned}
$$

Appendix C. Stability analysis

The local stability of the equilibria we have found can be investigated by finding the leading eigenvalue of the (Jacobian) stability matrix of the model evaluated at each of the equilibrium values (for details of this technique see, e.g., Otto and Day, 2006). This tells us the effect that a small perturbation, ε, will have on the equilibrium values of a_2 and t_2 (ε is assumed to be small enough to ignore in second and higher order terms). If the leading eigenvalue is greater than 1 then the perturbation will grow over time and so the equilibrium is unstable, while if the leading eigenvalue is less than 1 the perturbation will shrink over time and the equilibrium is locally stable.

C.1. Polygynous model

To make algebraic analysis possible we again make the simplifying assumption that $t_2>0$ and that m is large, so that $P_{T_2}=1$. In this case for the A_2 fixation equilibrium, where $\{\hat{a}_2=1,\hat{t}_2=\delta\}$, the only non-zero eigenvalue is given by the following equation:

$$\lambda=\frac{f_{LL}\cdot(2-4\delta+2\delta^2)+f_{LH}\cdot(2\delta-2\delta^2)+f_{HL}\cdot(2\delta-s\delta-2\delta^2-s\delta^2)+2\delta^2-s\delta^2}{2\cdot[f_{LL}\cdot(1-2\delta+\delta^2)+f_{LH}\cdot(\delta-\delta^2)+f_{HL}\cdot(\delta-\delta^2-s\delta+s\delta^2)+s\delta-s\delta(1-\delta)^n+\delta^2-s\delta^2]}. \tag{C.1}$$

Eq. (C.1) is a rather complicated expression and we have been unable to find simple expressions of the parameters which allow us to see when the equilibrium will be stable. Instead, to investigate the effect of the various parameters on the eigenvalue we have varied one parameter at a time while holding the others constant, the results of this analysis are shown in Fig. 6. This figure demonstrates that $\lambda<1$ for a wide range of parameter values, the only time that $\lambda\geqslant 1$ is when δ is very low or when $s=0$, i.e. when the X_8 females show no preference for type 2 song. This suggests that this equilibrium is generally stable for plausible parameter values.

For the polymorphic equilibrium the only non-zero eigenvalue is a very complicated expression, but using a computer algebra system we have been able to perform the same graphical analysis as used above. The effects of each of the parameters is shown in Fig. 7. We see that the $\lambda>1$ for a wide range of parameter values which suggests that this equilibrium is generally unstable.

For the final equilibrium of the model where $\{\hat{a}_2=a_2,\hat{t}_2=0\}$ and we no longer assume that $P_{T_2}=1$, there are two non-zero eigenvalues: $\lambda_1=1$ and $\lambda_2=\delta a_2 m$. The first eigenvalue reflects the fact that when $t_2=0$ any perturbation in a_2 will persist over time but will neither shrink nor grow. The second eigenvalue characterises the stability of a perturbation in t_2. $\lambda_2>1$ when $a_2>1/\delta m$, so if a_2 drifts to a value larger than this a perturbation to t_2 will grow and A_2 will become subject to frequency-dependent selection.

C.2. Monogamous models

For the monogamous model where the X_8 females choose after the other females, the equilibria are the same as for the polygynous model and so the same stability analysis applies here, except that the parameter n is fixed at 1.

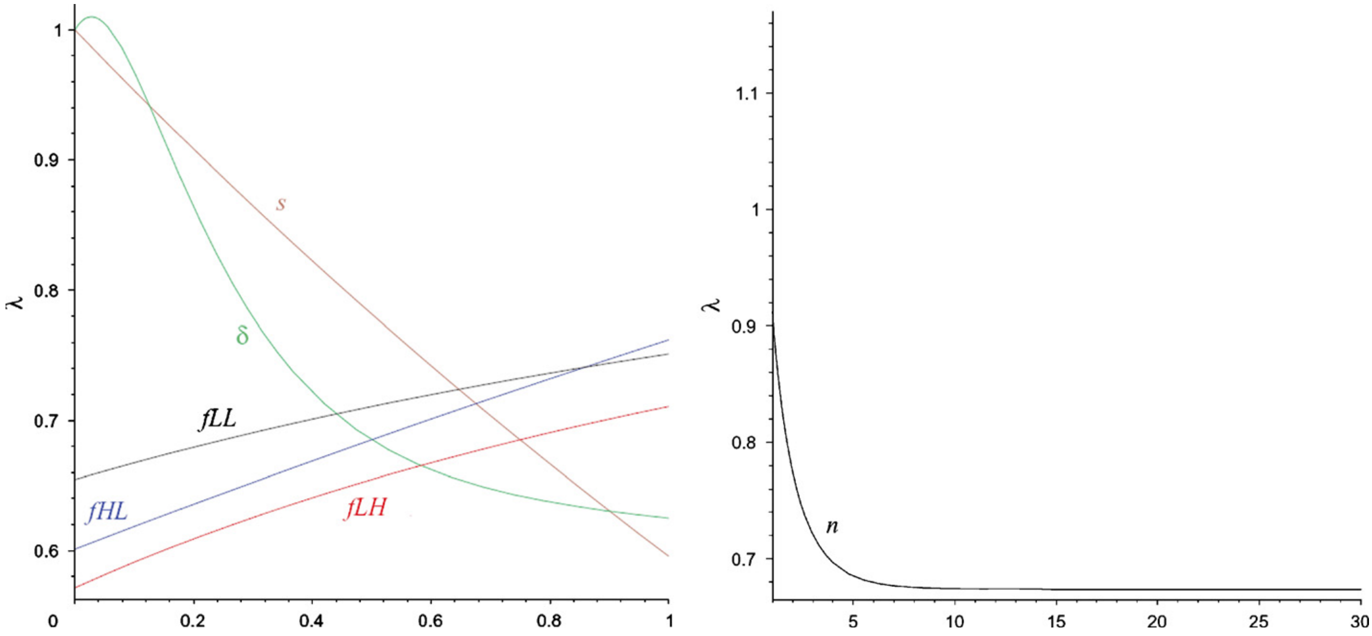


Fig. 6. Graphs showing the effect of the parameters on the leading eigenvalue of the A_2 fixation equilibrium for the polygynous model. The default parameter values used here are: $\delta = 0.5, m = 10, n = 5, f_{LH} = 0.75, f_{HL} = 0.5, f_{LL} = 0.25, s = 0.75$. Note that $\lambda < 1$ for all parameter values (except when $\delta < \sim 0.05$ and $s = 0$).

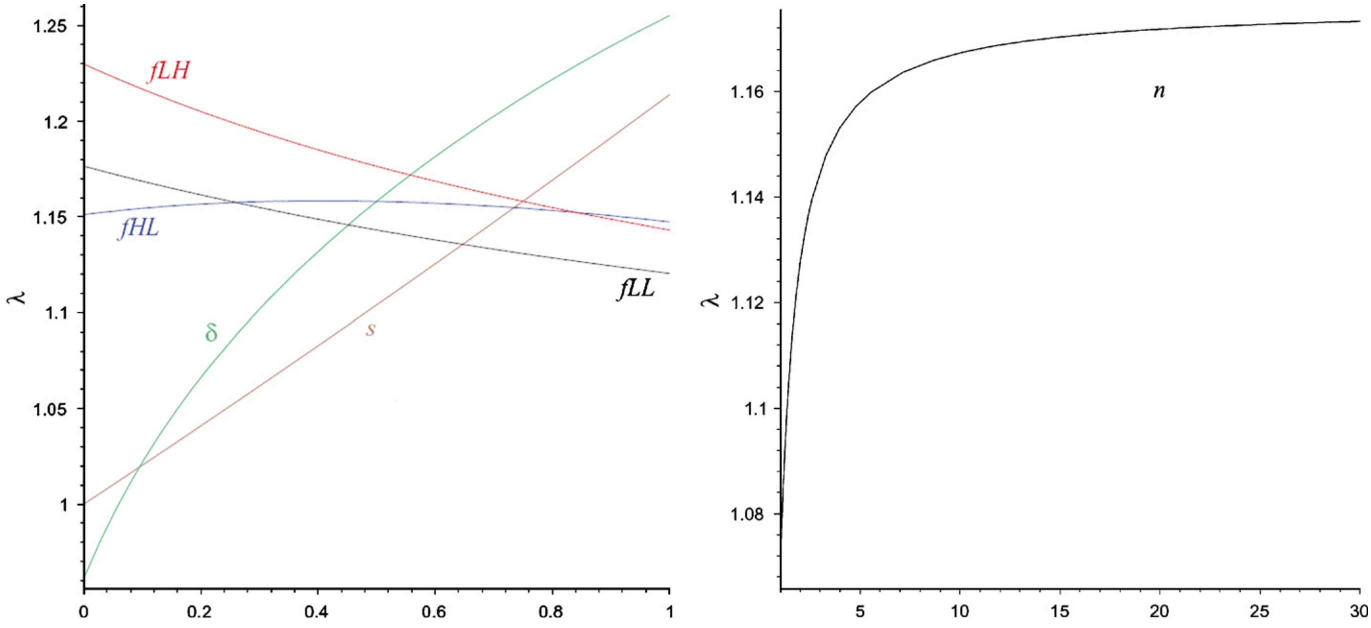


Fig. 7. Graphs showing the effect of the parameters on the leading eigenvalue of the polymorphic equilibrium for the polygynous model. The default parameter values used here are: $\delta = 0.5, m = 10, n = 5, f_{LH} = 0.75, s = 0.75, f_{HL} = 0.5, f_{LL} = 0.25$. Note that $\lambda > 1$ for a wide range of parameter values except when $\delta < \sim 0.05$ and $s = 0$.

For the case where the X_8 females choose before the other females there is no polymorphic equilibrium, but there is a fixation equilibrium where, again, $\{\hat{a}_2 = 1, \hat{t}_2 = \delta\}$. The only non-zero eigenvalue of this equilibrium is given by Eq. (C.2). This is again a rather complicated expression and so we provide a graphical analysis of the effect of different parameter values in Fig. 8. We see again that this equilibrium appears to be stable over a broad range of parameter values:

$$\lambda = \frac{f_{LL} \cdot (2 - 4\delta + 2\delta^2) + f_{LH} \cdot (2\delta - s\delta - 2\delta^2 + s\delta^2) + f_{HL} \cdot (2\delta - s\delta - 2\delta^2 - s\delta^2) + 2\delta^2 - 2s\delta^2}{2 \cdot [f_{LL} \cdot (1 - 2\delta + \delta^2) + f_{LH} \cdot (\delta - s\delta - \delta^2 + s\delta^2) + f_{HL} \cdot (\delta - s\delta - \delta^2 + s\delta^2) + s\delta + \delta^2 - 2s\delta^2]}. \tag{C.2}$$

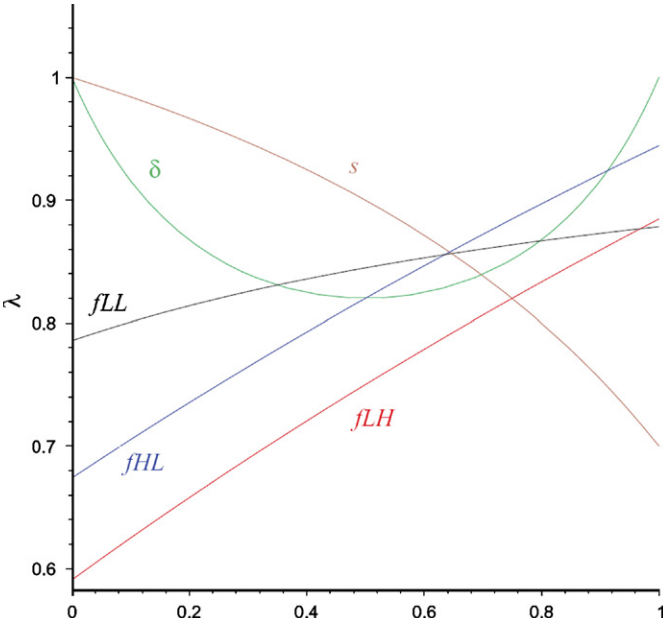


Fig. 8. Graph showing the effect of the parameters on the leading eigenvalue of the A_2 fixation equilibrium for the monogamous model when the X_8 females choose before the other females. The default parameter values used here are: $\delta = 0.5$, $m = 10$, $f_{LH} = 0.75$, $f_{HL} = 0.5$, $f_{LL} = 0.25$, $s = 0.75$. Note that $\lambda < 1$ for all parameter values tested here (except when $\delta = 0$ or 1 and when $s = 0$).

The eigenvalues for the $\{\hat{a}_2 = a_2, \hat{t}_2 = 0\}$ equilibrium of this model are the same as for the polygynous case and so the same analysis applies here. However, as there is no region of selection against A_2 in this model the result that $\lambda_2 > 1$ when $a_2 > 1/\delta m$ gives us the value of a_2 above which A_2 is positively selected for. Therefore, as δ and m increase this threshold value decreases and so the likelihood of invasion by A_2 increases.

References

Andersson, M., 1986. Evolution of condition-dependent sex ornaments and mating preferences: sexual selection based on viability differences. Evolution 40 (4), 804–816.

Aoki, K., 1989. A sexual selection model for the evolution of imitative learning of song in polygynous birds. Am. Nat. 134 (4), 599–612.

Ballentine, B., Hyman, J., Nowicki, S., 2004. Vocal performance influences female response to male bird song: an experimental test. Behav. Ecol. 14 (1), 163–168.

Brighton, H., Kirby, S., Smith, K., 2005. Cultural selection for learnability: three principles underlying the view that language adapts to be learnable. In: Tallerman, M. (Ed.), Language Origins: Perspectives on Evolution. Oxford University Press, Oxford (Chapter 13).

Buchanan, K.L., Spencer, K.A., Goldsmith, A.R., Catchpole, C.K., 2003. Song as an honest signal of past developmental stress in the European starling (*Sturnus vulgaris*). Proc. R. Soc. London Ser. B 270, 1149–1156.

Buchanan, K., Leitner, S., Spencer, K., Goldsmith, A., Catchpole, C., 2004. Developmental stress selectively affects the song control nucleus HVC in the zebra finch. Proc. R. Soc. London Ser. B 271 (1555), 2381–2386.

Catchpole, C.K., Slater, P.J.B., 1995. Bird Song: Biological Themes and Variations. Cambridge University Press, Cambridge.

Darwin, C., 1879. The Descent of Man, and Selection in Relation to Sex, second ed. John Murray, London (Reprinted in 2004 by Penguin).

Draganoiu, T.I., Nagle, L., Kreutzer, M., 2002. Directional female preference for an exaggerated male trait in canary (*Serinus canaria*) song. Proc. R. Soc. London Ser. B 269, 2525–2531.

Grafen, A., 1990. Biological signals as handicaps. J. Theor. Biol. 144, 517–546.

Hasselquist, D., 1998. Polygyny in great reed warblers: a long-term study of factors contributing to male fitness. Ecology 79 (7), 2376–2390.

Heywood, J.S., 1989. Sexual selection by the handicap mechanism. Evolution 43 (7), 1387–1397.

Kirkpatrick, M., 1982. Sexual selection and the evolution of female choice. Evolution 36 (1), 1–12.

Lachlan, R.F., Feldman, M.W., 2003. Evolution of cultural communication systems: the coevolution of cultural signals and genes encoding learning preferences. J. Evol. Biol. 16, 1084–1095.

Lachlan, R.F., Slater, P.J.B., 1999. The maintenance of vocal learning by gene–culture interaction: the cultural trap hypothesis. Proc. R. Soc. London Ser. B 266, 701–706.

Laland, K., 1994. Sexual selection with a culturally transmitted mating preference. Theor. Popul. Biol. 45, 1–15.

Lauay, C., Gerlach, N.M., Adkins-Regan, E., Devoogd, T.J., 2004. Female zebra finches require early song exposure to prefer high-quality song as adults. Anim. Behav. 68, 1249–1255.

Marler, P., Slabbekoorn, H. (Eds.), 2004. Nature's Music: The Science of Birdsong. Academic Press Inc., Ltd., London.

Naguib, M., Nemitz, A., Gil, D., 2006. Maternal developmental stress reduces reproductive success of female offspring in zebra finches. Proc. R. Soc. London Ser. B (Published online).

Nowicki, S., Searcy, W.A., 2004. Song function and the evolution of female preferences: why birds sing, why brains matter. Ann. NY Acad. Sci. 1016 (1), 704–723.

Nowicki, S., Peters, S., Podos, J., 1998. Song learning, early nutrition and sexual selection in songbirds. Am. Zool. 38, 179–190.

Nowicki, S., Hasselquist, D., Bensch, S., Peters, S., 2000. Nestling growth and song repertoire size in great reed warblers: evidence for song learning as an indicator mechanism in mate choice. Proc. R. Soc. London Ser. B 267, 2419–2424.

Nowicki, S., Searcy, W.A., Peters, S., 2002a. Brain development, song learning and mate choice in birds: a review and experimental test of the "nutritional stress hypothesis". J. Comp. Physiol. A Sens. Neural Behav. Physiol. 188 (11–12), 1003–1014.

Nowicki, S., Searcy, W.A., Peters, S., 2002b. Quality of song learning affects female response to male bird song. Proc. R. Soc. London Ser. B 269, 1949–1954.

O'Connor, R.J., 1984. The Growth and Development of Birds. Wiley, New York.

O'Donald, P., 1980. Genetic Models of Sexual Selection. Cambridge University Press, Cambridge.

Otto, S., Day, T., 2006. A Biologist's Guide to Mathematical Modeling in Ecology and Evolution. Princeton University Press, Princeton, NJ.

Oudeyer, P.-Y., 2005. How phonological structures can be culturally selected for learnability. Adaptive Behav. 13, 269–280.

Riebel, K., 2003. The "mute" sex revisited: vocal production and perception learning in female songbirds. Adv. Study Behav. 33, 49–86.

Slater, P.J.B., 2003. Fifty years of bird song research: a case study in animal behaviour. Anim. Behav. 65, 633–639.

Spencer, K.A., Buchanan, K.L., Goldsmith, A.R., Catchpole, C.K., 2004. Developmental stress, social rank and song complexity in the European starling (*Sturnus vulgaris*). Proc. R. Soc. London Ser. B 271, (Suppl.) S121–S123.

Spencer, K.A., Wimpenny, J.H., Buchanan, K.L., Lovell, P.G., Goldsmith, A.R., Catchpole, C.K., 2005. Developmental stress affects the attractiveness of male song and female choice in the zebra finch (*Taeniopygia guttata*). Behav. Ecol. Sociobiology 58, 423–428.