EVOLUTION OF SONG CULTURE IN THE ZEBRA FINCH

by

OLGA FEHER

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2009

This manuscript has been read and accepted for the Graduate Faculty in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Ofer Tchernichovski

________________ ____________________________________

Date Chair of Examining Committee

Dr. Laurel Eckhardt

________________ ____________________________________

Date Executive Officer

Dr. Joshua Wallman________

Dr. Carolyn Pytte__________

Dr. Marc Schmidt__________

Supervisory Committee

The City University of New York

Abstract

EVOLUTION OF SONG CULTURE IN THE ZEBRA FINCH

by

Olga Feher

Adviser: Dr. Ofer Tchernichovski

Cumulative cultural evolution is when behavior in subsequent generations of learners builds on the accumulated information of previous generations to such an extent that no individual learner can produce the behavior on its own. Many examples exist in humans, but in nonhuman animals there are only a handful of suspected cases. Here, we provide the first demonstration of cumulative cultural evolution in the laboratory in nonhuman animals. We raised zebra finches in complete acoustic and social isolation to create "uncultured" animals. Isolate zebra finches sing unstructured songs that are different from wild-type songs in many aspects such as spectral details of syllables and syntactic organization. We developed an automated procedure to quantify the differences between isolate and wild-type song at different timescales of song structure: spectral features, duration of song notes and song rhythm. We then used the isolate birds to teach their songs to juveniles who became the tutors for the next generation of learners and so on recursively. We followed the evolution of isolate song over multiple generations. We found that isolate song was gradually transforming into wild-type song over 3-4 learning generations. In addition to this experiment where we trained young birds in individual tutor-pupil pairs, we established a semi-natural colony with an isolate founder and tracked song changes over multiple generations of learners. In the colony, the song also

progressed towards wild-type song in a few generations, but some of the details of the changes differed between the two conditions. The rapid evolution indicates that wild-type song culture is encoded in every bird, but it takes multiple generations to surface. The young birds used imitation biases to change isolate song features into wild-type features.

Acknowledgements

I thank my intellectual father and adviser, Ofer Tchernichovski, for being so crazy in Bio 102 that I immediately wanted to work with him. Thank you, Ofer, for convincing me to stay here and letting me learn and do things I was interested in and for being so loyal that I never had to worry about your continuous support (except maybe once or twice). Thanks also for not only tolerating my cruel sense of humor but actually enjoying it! You really have been the best adviser I could have had. Thanks, Benjy for telling me in the coffee shop that I could do whatever I wanted and for standing by me in absolutely everything since then, and, most importantly, for being such a considerate and easy husband. I also thank our amazing daughter, Zsófi, for keeping me on my toes and giving me emotional fuel with her sweetness during the great times as well as the hard times. Very importantly, I have to thank my mom, Olgi, who has always made me feel special and, except for my gothic outfits at fourteen, never judged me or tried to change me. Her amazingly warm presence in my life gave me lots of courage to try new things, including this degree. I am also grateful for my wonderful siblings, Árpi, Kriszti and Böbi whom I deeply love and admire. I thank Noah Perlut who inadvertently made me realize that I wanted to be a Biologist. I thank the fantastic Neuroscience faculty at City College who have all been incredibly friendly and supportive. I have received lots of guidance and friendship from Josh Wallman in particular. Thanks, Josh. I thank my advising and examining committee for being fair and giving me lots to think about. Last, but not least, I would like to thank all my lab mates who are great lunch buddies and very nice people.

Table of Contents

List of Tables

List of Figures

Introduction

Culture and cumulative culture in animals

Culture has dozens of definitions and interpretations. Some ethologists consider any socially transmitted information culture (Bonner 1980), but in cultural evolutionary theory, it is widely accepted that, although there are instances of culture in nonhuman animals, many aspects of culture are restricted to humans. Before considering one of these, cumulative culture, that is central to this dissertation, we will discuss some examples of cultural variation in nonhuman animals. We will limit our discussion to behaviors that are socially transmitted, maintained over generations, and that produce population-specific variations.

Many examples of culture in animals relate to methods of acquiring food. For instance, arboreal black rats (*Rattus rattus*) in Israeli forests during a sensitive period in early life learn a particular and very efficient technique to open pine cones from their mother. Pine cone scales form a spiral – some spirals go left, others right. The technique involves finding the key scale at the bottom of the cone, identifying the direction of the spiral, and then removing each scale in turn to get to the pine nut which is hidden below it. This behavior does not appear spontaneously (Aisner & Terkel 1992). For example, naïve adult rats can open pine cones, but they do so by simply chewing a hole in its middle, and never revise their inefficient technique of obtaining the pines. Even their offspring do not improve the technique. It is assumed that at some point an especially creative rat must have come up with it on her own, but it is only maintained in the population because the technique is learned by new generations.

Other examples of culture in animals involve the usage of tools to get to food more efficiently. For example, chimpanzees stick probes into termite nests and manipulate them in a way that results in many termites getting caught on the stick (Goodall 1964). Young chimpanzees observe older individuals performing this trick and then practice it on their own. Moreover, in the Congo Basin, chimpanzees use an improved design, brush-tipped fishing probes that are different from the simpler probe sticks of East and West Africa (Sanz et al. 2009). Even more impressively, New Caledonian crows manufacture tapered tools from narrow leaves of pandanus trees which they use to extract invertebrates from the rainforest vegetation (Hunt 1996). Although tool use and manufacture is common among crows, only those in New Caledonia (*Corvus moneduloides*) make the tapered tools. In different regions of the main island, Grande Terre, there are differences in the design of the tools, with more northerly regions containing more sophisticated tools with more steps. The geographical distributions of the different designs overlap and are independent of ecological correlates, which implies that the designs are culturally transmitted and maintained, although there is no direct evidence for this (Hunt & Gray 2003). Building increasingly sophisticated tools by improving on previous designs is indicative of cumulative cultural evolution.

Cumulative culture, although widespread in humans, is extremely rare even in birds and mammals. An obvious human example is technological evolution: as an instrument-building behavior is transmitted across generations, instruments become more and more sophisticated, each time building on the previous state, so that it reaches a point when no single individual, no matter how clever, could invent the instrument from scratch. According to Boyd & Richerson (2005), cumulative cultural evolution assumes a gradual change in the behavior over many generations and a branching pattern of descent by modification, that is, populations that are more closely related share more characteristics than populations more distantly related. This, of course, is somewhat complicated by the horizontal transmission of culture, and since good documentation only exists for human culture (language and technological changes), investigation of cumulative cultural evolution in nonhuman animals may be challenging. According to Boyd & Richerson's criteria, one of the very few examples of nonhuman cumulative cultural evolution exist in songbirds, because regional dialects exhibit such geographical relationships (Marler & Tamura 1964). However, in the absence of empirical evidence, we cannot be certain that song dialects do not represent random fluctuations in singing behavior over generations and over space (analogous to genetic drift) as opposed to directional shifts that accumulate over generations (e.g., songs becoming more and more structured over generations), which would be more similar to what we see in human cumulative cultures. In this study, we provide the first evidence that wild-type birdsong evolves in a cumulative manner and results in behavioral variants that cannot be produced by an individual animal.

Song culture in birds

Vocal learning

The complex behavior that allows song cultures to develop is vocal learning. Songbirds, just like human infants, are born with a predisposition to learn complex vocalizations, which they do during an early stage in development by gradually modifying their vocal output to match that of adults (Immelmann 1969). This is, of course, similar to how humans learn to speak languages. Vocal learning behavior is rare: besides humans and oscine songbirds (Thorpe 1958), strong evidence only exists for parrots (Pepperberg 1999), hummingbirds (Baptista & Schuchmann 1990), and cetaceans (Reiss & McCowan 1993, Foote et al. 2006). There is some indication that bats may be vocal learners as well (Boughman 1998). Vocal learners are able to change their vocal behavior based on auditory stimuli, and they imitate adults to achieve species-typical vocal behavior. In addition, there are some characteristics of the developing vocalization that are shared among vocal learners: for example, sounds become more structured and less variable during development. In fact, the early song of young birds is sometimes referred to as babbling, because the vocal exploration involved in it is reminiscent of pre-linguistic vocalization in human infants (Doupe & Kuhl 1999).

The sensitive period

As mentioned above, vocal learners acquire their vocalizations by imitating their parents during an early stage in development. This period is called the sensitive period, the time during which the young bird is able to memorize adult sounds and by gradually changing his own vocalizations, match those sounds. In some species, like the song sparrow, the two processes (memorization and practice) are removed from each other in time, as song sparrows listen to adult males in the spring, but do not produce their own songs until the next spring, which they do even when isolated (Marler 1964). Therefore, vocal learning in birds requires the formation of a long-lasting memory of the model song, often referred to as the ‘auditory song template’ (Konishi 1965). Vocal learning is then the gradual process of mapping the template onto a set of motor commands which becomes the acoustic signal, the song. It is thought, although without firm supporting evidence, that the mechanism of sensory-motor conversion necessary for achieving vocal imitation might be similar in birds and humans.

Song dialects

Many oscine bird species can only acquire new songs during the early sensitive period. These birds are called “age-limited” learners, and an example is the zebra finch (*Taeniopygia guttata*). However, some species can also learn new songs as adults; we refer to these as “open-ended” learners, and this group includes the canary (*Serinus canaria*). Open-ended learners often learn new songs in a seasonal pattern (Nottebohm 1987). Errors in song learning and faithful return to natal areas are thought to give rise to geographical variation in the songs of most birdsong species. The variation in song between populations is much larger than within a population (Thompson & Baker 1993), which is the basis for regional song dialects. Marler and Tamura were the first to systematically study song dialects and discovered that over large geographical distances the songs become less and less similar (Marler & Tamura 1962). Cross-fostering experiments proved that these dialects are not genetic, rather, they are an outcome of cultural transmission (Marler & Tamura 1964). Shifts in song dialects can occur when the separation between neighboring populations is sharp, often in the form of a geographical boundary (Thielcke 1969, Baker 1987). However, sometimes sharp boundaries between dialect populations do not coincide with habitat changes (Baptista 1977). Several studies demonstrated that song dialects can remain stable over many generations, probably because of bounded cultural transmission within stable populations with relatively little migration (Trainer 1983). When modifications to the dialect arise, they spread through the population by social transmission, and as more and more individuals (including the new generations) have incorporated the new inventions, the modifications become a stable part of the local dialect. This way, song dialects can evolve dynamically and cumulatively within a population. Payne et al. (1981) recorded the songs of Village indigo buntings (*Vidua chalybeata*) in southern Michigan over a 15-year period. He found that although 80% of the young birds copied song types from neighbors when establishing territories, they also made changes and over the 15-year old study period, only a few song types persisted. However, the half-life of songs was roughly three times longer than the half-life of the individuals.

What drives the evolution of song dialects?

There are several social interactions that underlie the cultural evolution of song dialects. In some species, especially age-limited learners who grow up in the presence of their parents, juveniles often imitate from their fathers, but only slightly more than from other males (Williams 1990). Young males have been found to imitate adult males who interacted with them most, including feeding them (Mann & Slater 1994), and even from those who were more aggressive towards them (Clayton 1987). Migrating birds, who arrive at a geographical location to establish a territory, often copy neighbors (Payne 1981). Village indigo birds copied their songs more from successfully breeding males than from males who were unsuccessful (Payne 1985).

Females probably play an important role, as well. The mating preferences of female cowbirds can be measured due to their display of copulatory postures. They respond significantly more to songs of reproductively successful males (West et al. 1981). Females can influence the persistence of the local dialects, because they were shown to preferentially mate with males who sing familiar dialects (Baker and Cunningham 1985, Slabberkoorn & Smith 2002). There are ways in which females can contribute to song diversity and complexity. For example, song sparrow females have been found to show preference for males with larger repertoires (Reid et al. 2004). In addition, Okanoya’s studies revealed that female selection can result in higher song complexity (2004). He studied Bengalese finches, a domesticated strain of the wild white-rumped munia, that for the last 240 years has been maintained in captivity in Japan. The song of the munia is much simpler and more stereotyped, whereas the Bengalese finch’s songs are complex and variable in syntax. Okanoya and his colleagues measured estradiol levels and nesting activity in females after stimulation with either complex domesticated or simple wild-type song, and found that both measures were higher after stimulation with complex songs. They hypothesized that a sudden elimination of predation pressures and the females' preference for complex song resulted in the complex song syntax of Bengalese finches.

Documentation of changing cultures

Although cultures, cumulative or non-cumulative, have been studied intensely in both humans and animals, the development of existing cultures has only rarely been studied. In humans, Nicaraguan Sign Language (NSL) is perhaps the only well documented example. With the foundation of a school for special education in Nicaragua in 1977 (Kegl 2001), previously isolated deaf children from across the country came together for the first time. These children communicated in home signs within their families, which varied widely between families in form and complexity. Although the education was conducted in Spanish, the children began communicating with their peers, and within a few years established a sign language that bore grammatical properties similar to other existing sign languages. The spontaneous development of NSL is a fascinating example, but it leaves many questions open, such as what is needed for the language acquisition system to be activated, whether rich lexical or grammatical input is necessary and how exactly it happens. For obvious ethical reasons, it is impossible to subject human children to experimental manipulation to answer these questions.

In animals, culture has been studied mostly in the wild, where researchers must rely on small samples of the behavior. To our knowledge no studies have been done in a laboratory setting, where the environment can be controlled over generations, as culture develops. In addition, most studies rely on a qualitative assessment of cultural change. Besides a tight experimental control, there is a need for robust quantitative methods that allow us to measure cultural evolution reliably. This study hopes to overcome this obstacle by proposing a model system to study cultural evolution in the laboratory.

Goals and rationale of this dissertation

We developed a controlled experimental setting which recreates conditions that are conducive to cultural evolutionary changes, which we have recorded and analyzed. We take advantage of the distinction between isolate and wild-type zebra finch song. As mentioned before, zebra finches are closed-end learners, which means they learn only one song early in life and then lose the ability to learn more. There is also a strong sexual dimorphism in the organization of the brain (Nottebohm & Arnold 1976), which manifests itself in behavior, because only males produce song. Young males learn their songs from older males by imitation. Zebra finches, when raised in complete acoustic and social isolation, will invent a song and will stabilize on a stereotyped song structure upon maturity and sing that song for the rest of their lives. Such improvised songs are called isolate (ISO) songs. These songs differ in several aspects from those that wild birds or birds reared in laboratory colonies produce (Price 1979, Williams 1993), which we will refer to as wild-type (WT) songs.

WT songs can be regarded as the cultured form of singing. They are the result of many generations of cultural transmission. ISO songs, on the other hand, represent the innate behavior that emerges without any social transmission or imitating. Of course, even ISO song has a learned aspect to it, because the birds make changes to it based on the auditory feedback of hearing themselves sing (Price 1979). The evidence for this is that birds deafened in early development, and therefore deprived of auditory feedback, produce a virtually “structureless” song (Nottebohm 1966), which is much more abnormal than ISO song. In fact, ISO song goes through developmental changes as does WT song, and it becomes more structured and more stable during ontogeny. The level of structure and stability is, however, lower than in WT songs. In addition, the sensitive period is also extended in ISO birds, and they are able to learn new songs or modify their songs after WT birds have already crystallized and incapable of new imitation - beyond 100 days or so (Eales 1985).

From a cultural evolutionary perspective, one would expect that, given enough time, and in appropriate conditions, ISO songs should evolve into WT songs, assuming that the wild-type birdsong culture is constrained by innate components. Therefore, by training young birds with ISO song, we may be able to create a situation where there will be strong cultural evolutionary forces, such as those that restrict local dialects, that drive the song towards the WT. We test if ISO song evolves into WT song over multiple learning generations in different social situations: either one-to-one in an impoverished social setting or in a colony. In the one-to-one setting, we establish tutoring lineages starting from ISO birds, and document how the song changes across multiple learning generations. In the rich social setting, we establish an isolated semi-natural colony with an ISO bird, and follow the song as it is imitated by the offspring that are born into the colony.

Several possible mechanisms could account for a progression towards WT-like song. We present 3 non-mutually-exclusive scenarios.

Scenario 1. Cultural evolution of songs by sexual selection

Darwinian mechanisms have been proposed to be responsible for the transmission of cultural information (Campbell 1965), and even Darwin himself claimed that acquired variation was compatible with the forces of natural selection (1874, Descent of Man). Memetics, put forth by Richard Dawkins (1976), claims that the units of cultural transmission, called memes (in our case, song syllables), are self-replicating and therefore analogous to genes. Clearly, there are fundamental differences between cultural and genetic transmission, one being that culture can spread horizontally but genes can only proceed vertically. Horizontal transmission is a crucial element in the evolution of culture, because it allows young animals to learn from their peers or authoritative figures besides their parents. Examples abound in human culture, but there is evidence for it even in birdsong, since siblings can be influenced by each other’s imitations (Tchernichovski 1998). Nonetheless, cultural learning has been claimed to be a highly adaptive behavior because an efficient transfer of cultural information will likely make the receiver fitter

(Bonner 1980). In some cases, the adaptive advantage is obvious, as in having a more efficient method to gather food or being able to recognize other individuals in the same population. Sometimes, cultural inheritance can reduce fitness (Cavalli-Sforza and Feldman 1981) in individuals, but Henrich and Boyd (2002) maintain that at the population-level fitness-enhancing traits must be preserved in subsequent generations.

In this scenario, random "mutations" occur in the imitation of ISO song, females prefer songs that are more WT-like, and chicks imitate more from their fathers. In this case, in analogy to Darwinian evolution, the song syllables are like genes, females are like the "selecting environment", and the male songs that "survive" are transferred to the next generation of males, eventually optimizing female choices, namely towards WT features that the females prefer. In this scenario, the WT song is encoded in female preference, which provides the engine for cultural change, whereas males imitate randomly – WT-like song just provides them with a sexual selection advantage. If the transformation of ISO song into WT is by random modification and natural selection, we expect that it will only happen in the colony setting and it will take a long time.

Scenario 2. Cultural evolution of songs by selective imitation

Selective imitation means imitating a certain action only in particular circumstances, when there is good reason to do so. It assumes that the learner applies decision rules to the cultural variants he observes (Boyd & Richerson 2005). Children do this often during development, and recently the domestic dog has been shown to possess the ability to use judgment in imitation (Range et al. 2007). This mechanism is different from natural selection in that the subsequent song imitations will not happen by random mutations, rather, by active selection, so the birds bring innate predispositions to the learning process.

In this scenario, progression towards WT occurs directly by the learners during imitation. Due to their improvised nature, ISO songs vary across ISO birds. Moreover, ISO syllables within a particular bird's song can vary in "abnormality." This scenario suggests that pupils imitate selectively those tutor syllables or songs that are more WT-like. For example, if the original ISO song has 5 syllable types, one similar to WT syllables, his pupil will imitate that WT-like syllable and improvise the rest of his syllables (like an ISO bird). Such recursive process of imitating syllables that are more WT-like should eventually lead to WT songs. If these forces are at play in our experiment, we expect that any approximation of WT song will happen irrespective of the social environment, that is, in both the colony setting and the one-to-one condition. In addition, only selected syllables will be copied and improvisation rate in subsequent learning generations will be high. The progress toward WT will take many generations of learners as WT-like syllables need to be selected in every generation from improvised, ISO-like songs, which slows the rate of accumulation.

Scenario 3. Cultural evolution by biased imitation

Cultural evolutionary theory does not distinguish between selective imitation and biased imitation, because one of the prerequisites of cumulative cultural evolution is that the cultural information itself is not modified from generation to generation, but is faithfully inherited (Bonner 1980). Therefore, the question is not how to change existing cultural information, but how to copy from what is available to be more successful. Humans use innate or learned biases to selectively imitate certain individuals, for example those who are more successful or those who are imitated more frequently by their peers (Boyd and Richerson 2005). In our case, it is important to distinguish between the two, as they involve completely different strategies and have different implications, therefore biases are used in a different context here. By biases, we mean imitation biases that have a direct phenotypic effect on the learners' songs, that is, the model song is altered by the learner during imitation.

In this scenario, as in Scenario 2, the learners guide the cultural evolutionary process. The difference is that in biased imitation pupils imitate both ISO-like and WT-like syllables, but change them actively during the imitation process. Syllables are stable and inherited but with modifications that steer the song culture toward WT-like songs. The ISO features change gradually into WT features as each generation of learners contributes to the process.

Biased imitation produces nonrandom modifications which accelerate the rate of evolution (Richerson and Boyd 2005). The birds are active participants and contributors to evolutionary changes. No improvisation is involved, as birds modify existing syllables. We may experience a slowing down in the progression towards WT which may cause some features to asymptote as the approximation of WT is more complete. This is because the innate biases the birds will employ to change the ISO-like songs will have a smaller effect when the deviation from WT song is smaller. We expect that if imitation biases act on song imitation, young males will not need the help of females, so we will see the effect in the impoverished social setting as well as in the colony. The difference between Scenario 2 and 3 is that in this scenario, all the syllables will be imitated but altered rather than some WT-like syllables being more preferentially imitated. Abnormal syllables should change more than WT-like syllables but both kinds will appear in pupils' songs.

Depending on the outcome of our experiment, we will be able to infer some of the cultural evolutionary mechanisms involved in the transfer of cultural information.

Chapter 1. How isolate songs differ from wild-type songs

Background & Rationale

Like many other oscine songbird species, zebra finches require exposure to adult song to develop a species-typical, wild-type (WT) song. There is a relatively short sensitive period, from about day 30 post-hatch to day 90-100, during which the young male zebra finch gradually learns his song. He does this by constantly changing his vocal output in reference to external (tutor) songs, using auditory feedback. By the end of this period, the young bird’s song is structured and stable, and it will remain like that for the whole lifetime of the bird.

In the wild, as well as in various laboratory settings, songs share several species-typical features. When we examine songs developed in a semi-natural colony setting, where young birds are surrounded by siblings, females, and a number of related or possibly unrelated males (all potential tutors), and compare these songs to those developed in a socially impoverished setting, where a juvenile male is housed with a single adult male, we see a similar outcome of song development and of imitation. Even when young birds learn songs from playbacks, in complete social isolation, song development is within the norm, although sometimes less accurate than that observed in live-tutored birds. In contrast to these conditions, zebra finches raised in complete social *and* acoustic isolation are known to produce abnormal or impoverished songs (Thorpe 1958, Marler & Tamura 1964), but very few studies looked at the differences between WT and ISO songs quantitatively.

The earliest study to investigate systematic differences between WT and ISO song was conducted by Price (1979), who raised zebra finches in different social conditions and then recorded their adult vocalizations. To produce ISO songs, he raised young males in either individual or group isolation without exposure to adult song. Group-isolated birds interacted socially and acoustically with their peers. Individually isolated birds only had acoustic access to other individually isolated birds of similar ages, but a separation wall prevented any social interaction. Although Price couldn't find any indication that they were imitating each other, it seems highly unlikely that the acoustical contact had no influence on their vocal developments and their final songs, because later studies found convergence of song in group-raised zebra finches (Volman 1995). Price's study found that isolates have fewer distinct bout notes, fewer types of call notes, and their note frequencies and durations were higher and more variable. However, the note stability and the general temporal patterning of their song bouts (introductory notes followed by song notes) were quite normal. In addition, long and short calls were less common in songs of isolates.

Thirteen years later, another study undertook a comparison between ISO and WT songs, this time with regards to reproductive success (Williams 1992). The song characteristics were not quantitatively described in this study, and once again, subjective criteria were used to assess whether a song looked ISO-like or WT-like (the authors used the terms untutored-like and tutored-like). Some of these criteria included the prevalence of higher frequency syllables, upsweeps and syllable repetitions in ISO song. Upsweeps and syllable repetitions are virtually never found in WT zebra finch song. ISO songs were found to consist of both ISO-like and WT-like syllables. Overall, ISO males had lower reproductive success, but, strangely, they were more prominent song models to juvenile males. Not only did their sons copy more from them than sons of WT males from their fathers, but unrelated juveniles also imitated their syllables more, and most interestingly, they preferred to imitate the ISO-like syllables and not the normal ones.

The descriptions and analyses by Price and Williams and colleagues were done before more sophisticated and quantitative sound analysis methods were developed, and have limited bearing on the problem of characterizing ISO versus WT songs. Although, there has been some intuitive understanding in the zebra finch field as to what makes ISO song different from normal WT song, we felt that for a clear understanding of the transition from ISO to WT songs, a new descriptive and quantitative study of WT and ISO songs at different timescales of song production is necessary. We set out to fill this gap at first by examining how spectral song features (e. g., pitch, frequency modulation et cetera) of WT versus ISO songs distribute. We then continued with features that capture longer timescales of song structure – duration of sounds and rhythms. We aimed at developing a metric system that could allow us to assess the "distance" between WT and ISO songs in a general manner, and in such a way that would allow us to test for approximation toward WT songs in subsequent experiments.

Looking at a variety of WT songs, even within the same colony, it is striking how different and diverse WT songs are across birds. In contrast to the intuition one could develop based on the literature, ISO songs are also extremely diverse across birds. Subjectively, some are quite like WT songs, and others are extremely abnormal. We will start by a subjective comparison of ISO and WT songs and will then address the question if, despite the high diversity, ISO and WT songs can be objectively classified into distinct categories, and if yes, in what song features and at what timescales of song structure are ISO and WT songs most separated? Finally, we will examine the cause for the differences between ISO and WT songs, by testing two non-mutually-exclusive hypotheses: that differences are due to the lack of song tutoring and that they are due to social isolation.

Methods

Experimental groups

Wild-type (WT) songs (n=52): To obtain a baseline for WT songs produced by domesticated zebra finches (*Taenyopygia guttata*) in typical aviaries, we used recordings from birds raised in two well-established colonies: The Rockefeller University Field Research Center colony and the Hunter College zebra finch colony. Both colonies have existed for over 20 years. Birds were kept in breeding rooms including family cages (with 6-12 birds) or in larger semi-natural aviaries. All birds were raised in a social environment including the mixed company of males and females. We used 52 songs altogether. We thank Fernando Nottebohm, Noam Leader and Cheryl Harding for making the WT songs accessible to us.

Isolates (n=17): Birds kept visually and acoustically separated from other birds during the sensitive period for song learning are called isolates. We raised 17 zebra finches from the City College of New York breeding colony in this way to create a large sample of ISO songs. All birds were raised by their parents in a dedicated cage until day 7 post-hatch (we start counting from the hatching of the first egg). Zebra finches do not imitate songs heard prior to day 20 post hatch (Roper & Zann 2006). See Appendix I for testing the effect of early exposure on songs. The father was then removed on day 7, and the cage (with the nest box) was taken to a nursery area housing mothers (who do not sing) and chicks only. Birds were raised by their mother, and on day 30, when the young can already feed themselves (and just at the onset of subsong), birds were placed in sound attenuation chambers and kept in complete isolation from day 30 until day 120 post-hatch or later. Colony management and isolation procedures have been described previously (Tchernichovski et al. 1999).

Visualization of songs

To observe the details of the sound frequencies we use spectral derivatives (Tchernichovski et al. 2000). Traditional sonograms represent the power of sound in a time-frequency plan, while the spectral derivatives represent the change of power. For each point of the two-dimensional time-frequency plan of a sonogram, one can measure the change of power from left to right (on time), from bottom to top (on frequency) or at any arbitrary direction. So spectral derivatives are derivatives of the spectrogram in an ‘appropriate’ direction in the time-frequency plan. The derivatives of each point are calculated in an angle that is perpendicular to the direction of frequency modulation. Spectral derivatives can be estimated using multi-taper spectral methods; they have the same resolution as spectrograms and are not artificially broadened.

Sound Analysis

All the analysis was performed using *Matlab 7*, except for feature calculations (below), which were done using *Sound Analysis Pro 2* (SAP II).

In our analysis we rely on the calculation of articulation based song features to describe zebra finch songs. Many of the previous attempts to automate the analysis of sound similarity used sound-spectrographic cross-correlation as a way to measure the similarity between syllables. The correlation between the spectrograms of two sounds was examined by sliding one on top of the other and choosing the best match (the correlation peak). However, measures based on the full spectrogram suffer from a fundamental weakness: the high dimensionality of the basic features. For example, cross-correlations between songs can be useful if the song is first partitioned into its notes and if the notes compared are simple. However, even in this case, mismatch of a single feature can reduce the correlation to baseline level. For example, a moderate difference between the fundamental frequencies of two complex sounds that are otherwise very similar would prevent us from overlapping their spectrogram images (a vertical translation will not help since the harmonics will not match).

The cross-correlation approach, as mentioned above, requires, as a first step, that a song be partitioned into its component notes or syllables. This, in itself, can be a problem. Partitioning a song into syllables or notes is relatively straightforward in a species such as the canary in which syllables are always preceded and followed by a silent interval. Partitioning a song into syllables is more difficult in the zebra finch, whose song includes many changes in frequency modulation and in which diverse sounds often follow each other without intervening silent intervals. Thus, the problems of partitioning sounds into their component notes and then dealing with the complex acoustic structure of these notes compound each other. The analytic approach of SAP II addresses both of the above difficulties. It achieves this by reducing complex sounds to an array of simple features and by implementing an algorithm that does not require that a song be partitioned into its component notes.

The analytic framework of SAP II is rooted in a robust spectral analysis technique that is used to identify acoustic features that have good articulatory correlates. The acoustic features that we chose to characterize zebra finch song are represented by a set of simple, one-dimensional measures designed to summarize the multidimensional information present in a spectrogram. A procedure for measuring similarity, based on such an analytic framework has important advantages. First, it enables the examination of one acoustic feature at a time, instead of having to cope with the entire complexity of the song of two birds. A distributed and then integrated assessment of similarity across different features promotes stability of scoring. Second, it is analytic: it evaluates each feature separately and tells you not only that two sounds are similar or different, but also in what sense they are similar/different.

For every 10 ms time window of song, we compute a number of song features to describe spectral changes in song (Fig. 1.1). In this dissertation, we focus on three features: amplitude modulation (AM), frequency modulation (FM) and goodness of pitch. These features have been described previously (Tchernichovski 2000), so we provide only a brief overview here. Amplitude modulation captures changes in the amplitude envelope of sounds. It is positive in the beginning and negative at the end of each sound.

Amplitude modulation is the overall time-derivative power across all frequencies within a range. Units of AM are 1/t, Sound Analysis Pro 2 does not scale AM, and time units are defined by the ‘advance window’ parameter. Frequency modulation is the angular component of squared time and frequency derivatives. This measure gives an absolute (unsigned) estimate of frequency modulation. The units are in radians. Goodness of pitch is the peak of the derivative-cepstrum calculated for harmonic pitch. Units are comparable to AM and can be converted to dB by subtracting a baseline and converting to log scale. All the features presented here (except for amplitude modulation) are *amplitude invariant*. That is, the amplitude of the sound recorded does not affect them and hence the distance between the bird and the microphone should have only a minor effect as long as the shape of the sound wave has not been distorted by the environment.

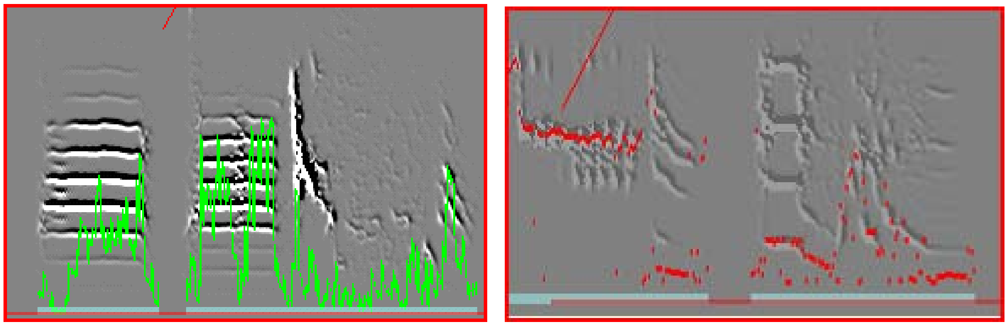


Figure 1.1 | Time series of spectral features overlaid on song sonogram. a. Goodness of pitch (green curve). **b.** Pitch (red curve).

The song features summarize the complex spectral image. A 1 second long sonogram includes 1000 spectral frames, each one represented by, let us say, 500 different frequency values. Therefore, the image includes 1000x500=5 million variables. If we use only 5 features to describe each spectral frame (for example, pitch, frequency modulation, amplitude modulation, Wiener entropy and goodness of pitch), the number of variables decreases to 5x1000=5000. However, this is still a lot, and 1 s is too short a sample to describe the singing behavior of the bird (we usually use at least 20s of song for each bird).

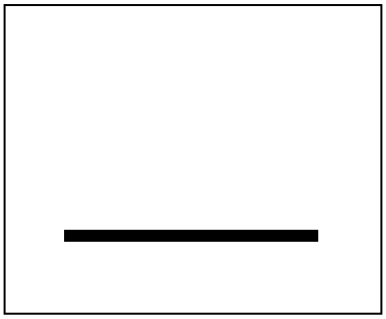
The next stage in simplifying our description is to ignore the time course of each feature (which can be very complicated as seen in the figure above) and look at how the features distribute. The distribution of song features is usually presented as a histogram with bins. The time course of 20,000 pitch values (of our 20 s song sample) can now be reduced to a histogram of 200 frequency values (bins). This decreases the number of variables from 20,000 to 200. However, this is still too many.

The next stage is summarizing the shape of the histogram. In statistics, we often describe distributions using parametric approaches. The most widely used approach is the Gaussian distribution, which is common in nature, and can be fully described by two parameters: mean and variance. This can potentially summarize the 200 frequency values in two measures. However, this approach is only meaningful if the feature distribution is Gaussian, and, unfortunately, this is not the case. Our feature distribution is skewed and often bimodal. In such cases, using mean and variance is rather meaningless – it would tell us very little about the shape of the histogram. Therefore, we are left with the histogram – the so called "non-parametric" description of the distribution (non-parametric is a poor terminology, because what it really means is multi-parametric).

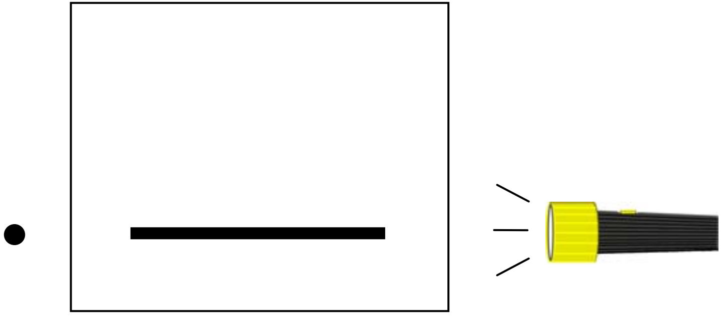
The next question is how to summarize the shape of the histogram without making strong (and often wrong) assumptions about how to model it? For such purposes statisticians developed so called "multivariate statistics" methods, which can be used to compress high-dimensional shapes and images. The most common method is called Principal Components Analysis (PCA).

In general, all the compressions we did so lose information – we cannot go from feature time courses back to the sonograms, nor can we go from the feature histograms back to the time courses. However, we are doing the "compressions" with some belief that the cost, in terms of losing information, is less than the gain: simplicity. Using PCA, we try to simplify our descriptions further at the smallest possible loss of information.

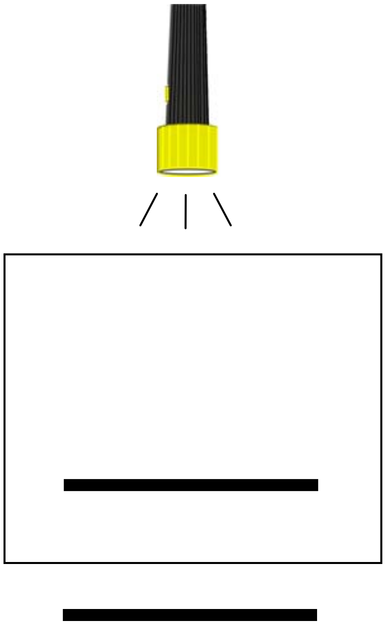
Let us consider a cartoon case – say that we have a piece of paper with a line drawn on it, and our task is to capture this line in a one dimensional representation by projecting it.



We can project the line by placing a flashlight on the right and look at the image on the left.

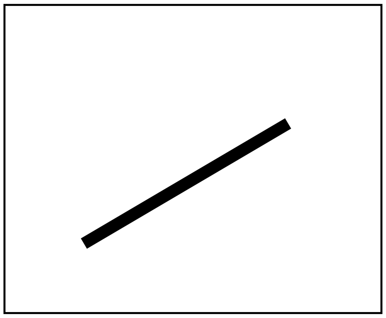


This projection simplified our line greatly, however, we lost most of the information, since the line turned into a single point. The solution is simple: place the light source above the surface of the paper.



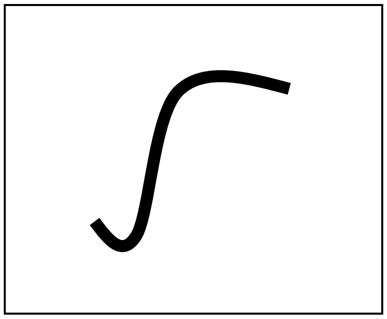
This time we did not lose any information, and we reduced a two-dimensional description (sheet of paper containing the line) into a one-dimensional representation of the object we were interested in describing.

Let us now consider another image.

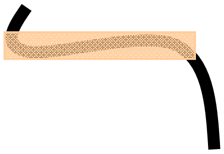


Clearly, we can achieve the same compression as we did with the horizontal line, if we rotate the image in reference to the flashlight. How can we do this automatically? To calculate the exact position automatically, we have to rotate the image slowly and choose the rotation that gives the longest line. That is, the one that maximizes the variance between the points of the line. Mathematically, rotation and projection means that we scale each axis. For example, in the first case of the horizontal line, we simply get rid of the Y axis, namely, multiply X values by 1, and multiply Y values by zero. In the case of a diagonal line, we multiply each axis by a number larger than -1 and smaller than 1. These numbers are called the loading factors or simply coefficients in PCA terminology. They give us all the possible rotations.

Finally, let us consider the following case.



This should be impossible, because there is no way we can capture this shape in one dimension. Nevertheless, some projections reveal more of it and are therefore better than others. We have to find the projection that maximizes the variance. However, there is still lots of information not captured by the projection. We now ask how much of the variance is not yet captured. By removing that “straight line” component of our curve (highlighted), we can now image a much simpler shape, with less structure in it:



Now, we repeat our procedure to capture the most information in the remaining parts of the shape: we rotate to maximize variance and project. Then we repeat again and again. The first rotation is called PC1, the second one PC2 et cetera. When we have rotated enough times, we should have no residual variance left. Each additional PC explains less and less of the variance.

Turning back to our analysis, we need to describe a histogram with 200 bins using PCA. The first rotation is a vector of 200 coefficients. If we multiply each frequency bin value by its coefficient and add them, we end up with one number, this is PC1. Now we go to the second rotation and do the same thing to get PC2.

The critical question is how much we have gained. If most of the variance can be explained by 2 PCs (PC1 and PC2), then we have, in terms of complexity of description, simplified our variables to the same degree as parametric analysis would have. However, in contrast to mean and variance, PC1 and PC2 give an excellent compression without any prior assumptions about the distribution of the data. The more structured the shape is – the better the PCA will work. In an image that is built by white noise and contains no structure, PCA will give no compression at all. So, PC1 and PC2 will explain very little of the variance, and we achieve nothing. PCA works well with high-dimensional data that contains lots of structure but cannot be simply described.

A serious shortcoming of the PCA technique is that interpretation of the results is difficult, because the PCs are just a collection of coefficients that do not mean anything in terms of what the data represents. This is why we find it so important to describe our songs with probability distributions and sonograms first, so we get an intuition about the data. We use PCA to visualize differences between ISO and WT songs and to provide a simple representation that is a good description of the feature distributions on which we can perform statistics.

Results

1.1 The structure of WT zebra finch song

When one listens to isolate zebra finch songs, it is often obvious that these songs differ from the songs of tutored birds. Although ISO songs are extremely variable, they often sound less rhythmic, less structured, more scratchy, longer in duration and higher pitched than WT songs. We will start by describing WT zebra finch songs, and then show common deviations we observe in ISO songs.

Continuous sounds in the zebra finch song are called song syllables. A syllable may be uniform in its spectral structure (Fig. 1.2a), or it may contain several distinct (uniform) sounds, called song notes (Fig 1.2b). Notes are often difficult to identify objectively as they blend into each other and the acoustic transition between them can be gradual rather than abrupt. Although an adult zebra finch song is highly stable, notes sometimes have some variability across renditions.

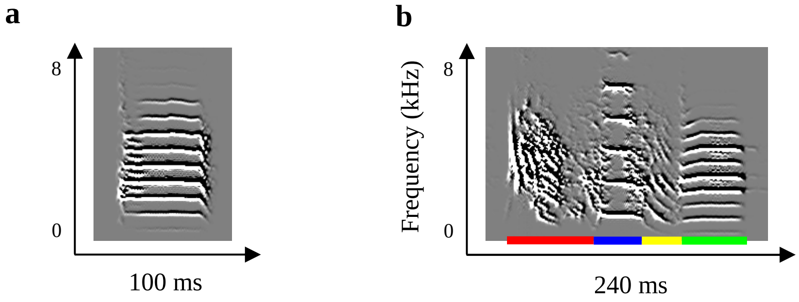


Figure 1.2 | Zebra finch song syllables. a. A syllable consisting of one song note. **b.** A syllable which consists of at least 4 acoustically distinct notes (underlined in different colors).

Syllables usually repeat in a fixed order composing a higher level of song structure called the song motif. A typical motif consists of 2-8 syllables that are separated by silence intervals. Figure 1.3 shows a motif that consists of 3 syllables (underlined in different colors).

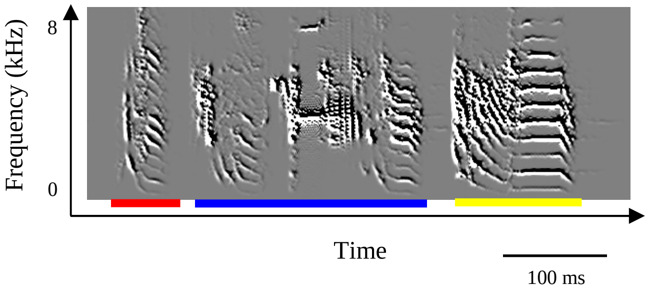


Figure 1.3 | Zebra finch song motif. The underlined syllables make up the motif, the smallest repeating unit of song.

Zebra finches sing in bouts: they first produce a number of introductory notes (Fig. 1.4, underlined in green), which are short, uniform, frequency-modulated calls, and then they repeat their song motifs, usually 2-7 times without stopping. Most zebra finches produce a single, unvaried motif as adults, and their song bouts are highly stereotyped repetitions of these.

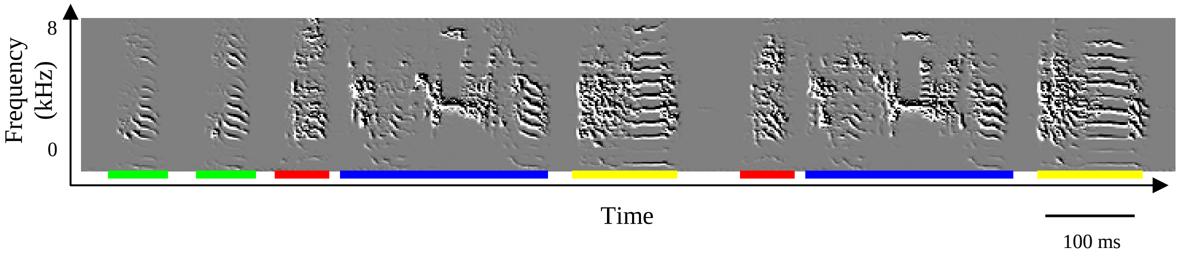


Figure 1.4 | Zebra finch song bout. A bout consists of introductory notes (green) followed by several repetitions of the song motif.

1.2 Subjective comparison of ISO versus WT songs

Some of the diversity of WT zebra finch songs can be seen in the examples presented in Figure 1.5. As shown, the motifs of WT zebra finch songs are highly stereotyped with very little variability between renditions. There are often rapid transitions between notes within syllables, with most notes being short (10-100 ms) and acoustically distinct. Longer syllables are usually complex, contain several notes, but even these syllables rarely exceed 200ms in duration. Most notes have harmonic structure, usually with distinct pitch (namely, with only one set of harmonic frequencies). Short (Fig. 1.5 Bird 3, green) and medium-length harmonics (Fig. 1.5 Bird 2, yellow) are common with either low (Fig. 1.5 Bird 1, first note of green syllable) or high pitch (Bird 3, near the end of yellow syllable). Many sounds are modulated, with decreasing frequency modulation (down-sweeps, Bird 3, pink) but up-sweeps are extremely rare. Amplitude modulations are strong and fast, and we often see series of extremely rapid amplitude modulations (vibratos, Bird 2 middle of green syllable). Short, high-pitched notes (Bird 4, red) are also common, usually, but not exclusively, within syllables. Long high-pitch notes are very rare.

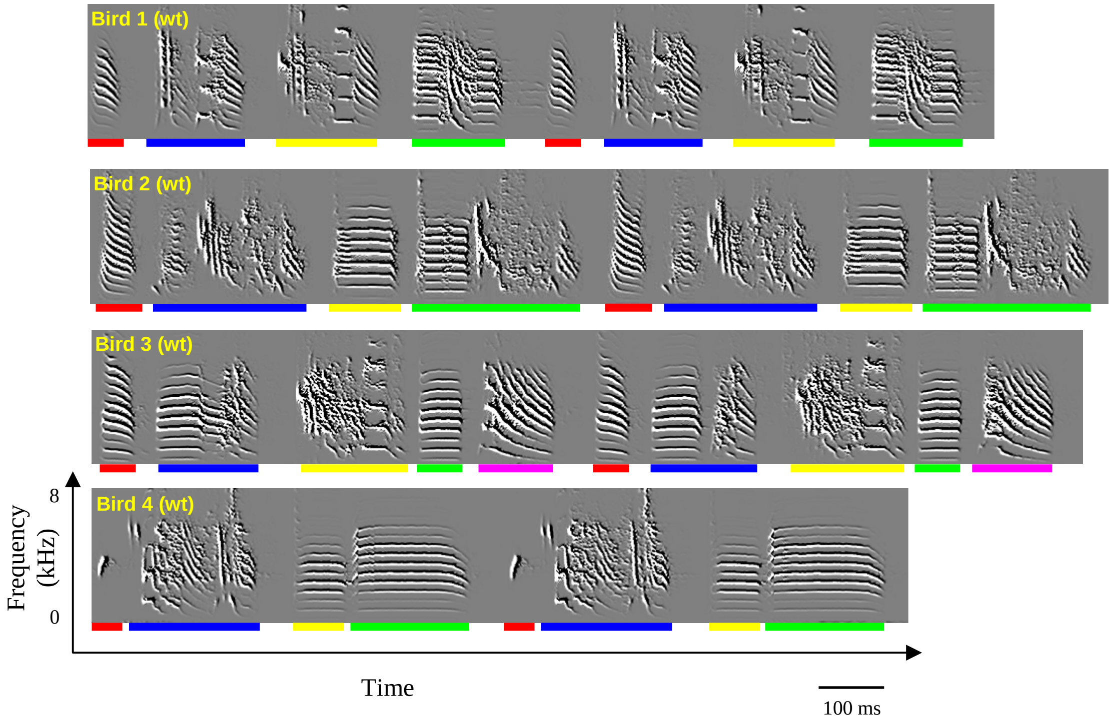


Figure 1.5 | WT zebra finch songs from 4 different birds. Each song example contains two motifs. Syllables are underlined in different colors (red, blue, yellow, green pink in order). WT songs are stable and syllables are always repeated in the same order with very little acoustic variability between renditions. There are fast transitions between notes (Bird 1, yellow syllable), many short and medium-length harmonics (e.g., Bird 4, yellow syllable) and frequency tends to be modulated downwards (e.g., Bird 3, pink syllable).

ISO songs are also diverse across birds, perhaps even more so than WT songs. In Figure 1.6, we present examples that demonstrate the diversity of ISO songs. Some seem as normal as WT songs while others have extremely aberrant song features. In some cases (Fig. 1.6 Bird 1) we see a rather WT-like acoustic structure, but bout organization is unusual. In other cases (Fig. 1.6 Bird 3) the bout structure looks WT-like, but spectral features are abnormal. Bird 1 sings a song that is quite complex with rapid note transitions and normal spectral features, however, there is only one complex syllable that is repeated over and over, and then the bout ends with three harmonic syllables, two of them short and one long. This type of bout organization is unusual and the harmonic syllable is longer than in WT songs. In Bird 2 the song begins with an abnormally long call-like syllable, followed by the stuttering of a single noisy syllable. Bird 3 has a WT-like bout organization, the syllables are repeated sequentially in a stereotypical order, but the syllables themselves are very short and simple, with no or little frequency modulation. Essentially, this bird sings non-modulated harmonic sounds of variable pitches. Bird 4 has a rather normal bout organization, but the most abundant syllable (underlined in blue) is highly atypical, consisting of a high-pitched, long harmonic and a broad-band scratchy sound. Bird 5 sings very simple and long harmonic syllables, and the final syllable (pink) has unstable harmonic structure that makes it sound scratchy. This syllable resembles a juvenile begging call in its spectral structure, and was likely derived from one. Bird 6 & 7 do not have a stable motif and they sing extremely long syllables (Bird 6, green & Bird 7, blue syllables), which are longer than a typical zebra finch motif. The long syllable of Bird 7 is so unusual that even to an untrained ear, it sounds more like a cry of pain than birdsong.

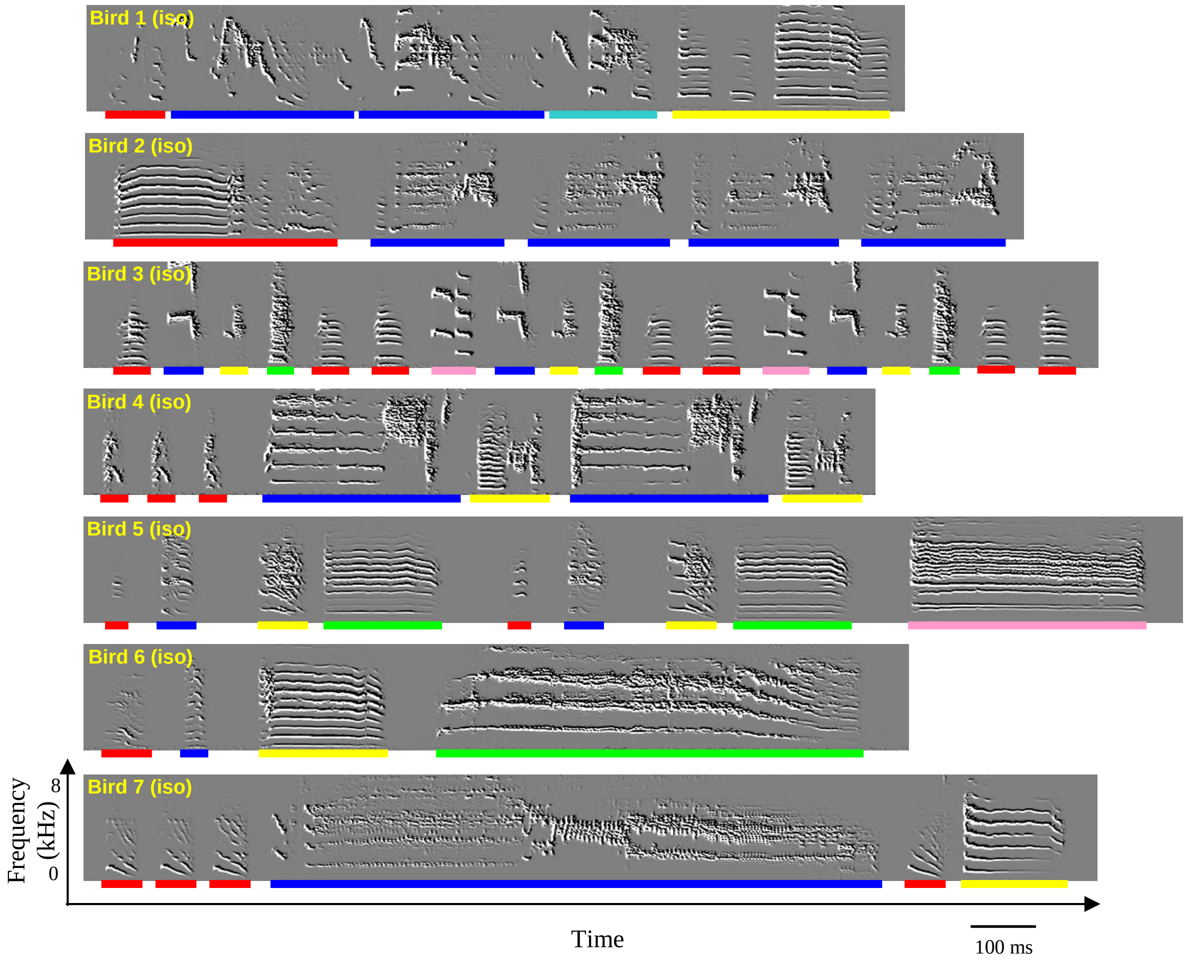


Figure 1.6 | ISO zebra finch songs from 7 different birds. Different colored lines indicate syllables (red, blue, yellow, green and pink in order). ISO songs often have internal repetitions of syllables (Bird 2, blue syllable) but no repeating motifs (an exception here is Bird 3), extremely long syllables (Bird 6, green syllable and Bird 7, blue syllable) and long harmonics (Bird 1, yellow syllable and Bird 5, green syllable).

If we compare typical syllables in ISO and WT songs, we notice that in ISO songs long harmonics dominate whereas in WT songs, short and medium-length harmonics are common. While in WT songs there are many down-modulated syllables, that is, sounds whose frequency contour slopes downwards, these are quite rare in ISO songs. Overall, the spectral complexity of ISO songs is much lower than WT songs. Next, we will compare song features between ISO and WT songs quantitatively at different timescales and levels of song production.

1.3 Comparing ISO and WT songs with respect to spectral frame features

As we noted above, spectral features such as frequency modulation seem to differ across WT and ISO birds. We tested for differences in the distribution of spectral features by calculating spectral features in 10ms time windows over samples of 20 song bouts per bird (about 10,000-20,000 independent samples per bird). We used *Sound Analysis Pro version 2* to calculate the following spectral features: pitch, frequency modulation (FM), amplitude modulation (AM), Wiener entropy & goodness of pitch. Spectral frame features describe the moment-to-moment vocalization of the bird.

Figure 1.7 presents the mean probability distribution histogram of all our WT (blue, n=52) and ISO (red, n=17) birds for one of the features that showed the strongest differences across WT and ISO: FM. Confidence intervals for the means are indicated by dashed lines. As shown, the distributions of FM show lower values in ISO songs. As we saw in Fig. 1.6, ISO songs contain more harmonic syllables, which developed from long and short calls, and less modulated song syllables. WT songs, on the other hand, have very few low FM sounds, and although in the middle range, there is not much difference, many WT songs have high FM. These are the vibratos and high sloped down-modulated sounds discussed previously.

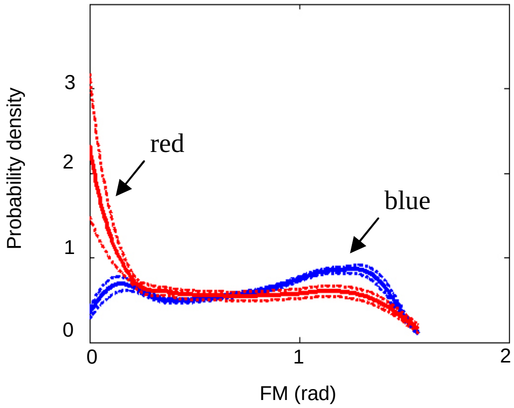


Figure 1.7 | Probability distribution histograms of WT and ISO birds. WT birds (blue line, n=52) show higher FM than ISO birds (red line, n=17). Dotted lines represent 95% confidence intervals.

We found that three features, FM, AM and goodness of pitch, provided the best separation between ISO and WT, so in our further analysis, we will focus on these three features. We calculated cumulative histograms from the probability distribution histograms (Fig. 1.8). The cumulative histograms, of course, showed the same differences between ISO and WT song features that the probability distributions showed, but visually they can show a better separation. ISO FM values are generally lower (Fig. 1.8a), indicating less modulation in sound frequencies. ISO AM values are closer to 0, which means that the changes in amplitude tend to be slower in ISO song (Fig. 1.8b). Goodness of pitch values tend to be higher in ISO song, which indicates that there are more or longer stable harmonic stacks corresponding to long calls (Fig. 1.8c). (See Chapter 1 Methods for information about units of AM and goodness of pitch.)

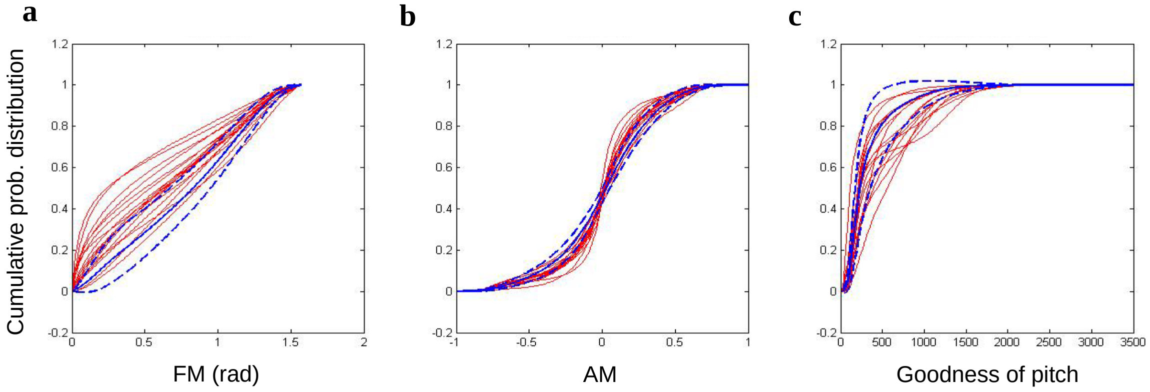


Figure 1.8 | Cumulative distribution histograms of WT and ISO birds for three spectral features: FM (a), AM (b) and goodness of pitch (c). Red lines (n=17) represent individual ISO birds, the blue line is the mean of the WT distribution (n=52), and the dashed lines represent the 95% confidence intervals for the mean of the WT data.

Distribution histograms are very useful to assess the general differences in WT and ISO songs for individual features, but the drawback is that every single feature for every single bird is represented by the long vector of the histogram values. To facilitate comparison across features, we summarized the feature distribution values for individual birds by using Principal Component Analysis (PCA). The input of the PCA is the cumulative histogram vectors of each feature of each bird. The principal components that are returned by PCA summarize the histograms of each individual bird in single numbers.

The first principal component (PC1) combines feature values so as to maximize variance across birds. This is equivalent to rotating the high dimensional histogram spaces of all birds and finding the projection that expresses most of the variance across birds. Then, the residual variance is rotated again to obtain the second PC (PC2) et cetera (more detailed description of PCA in Chapter 1 Methods).

Figure 1.9 shows the result of PCA based on three spectral features: FM, AM and goodness of pitch. Here, the first and second principal components (PC1 and PC2) are plotted against each other. As we can see, the red dots (ISO, n=17) and the blue dots (WT, n=52) separate from each other, or more precisely, occupy different parts of the continuum. There is some overlap, but almost all red dots are to the right of the blue dots. The distribution histograms of the PC1s of WT and ISO birds are shown under the graph, and they show a clear separation. The mean values of PC1 were significantly different between ISO and WT clusters ($p<0.001$, t-tests, $n_{WT}=52$, $n_{iso}=17$). The PCA does not provide additional information about the specific differences between the songs, but it is an objective means of confirming that there is indeed a real difference between ISO and WT songs.

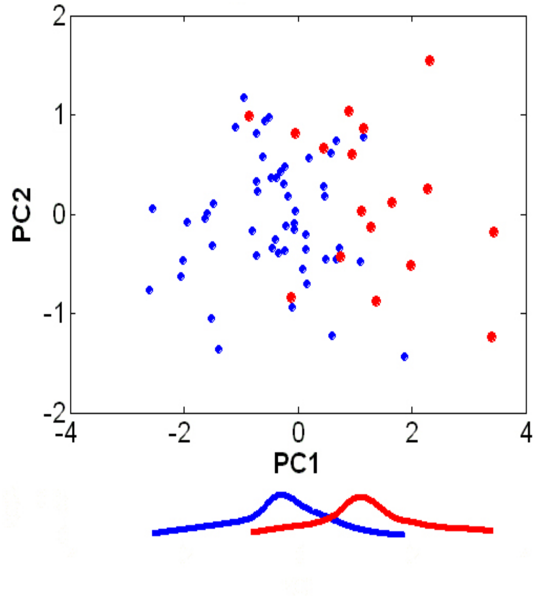


Figure 1.9 | First 2 principle components (PC1 and PC2) of spectral features of WT and ISO songs. Red dots (slightly larger) represent individual ISO birds (n=17), blue dots WT birds (n=52). Distribution histograms of WT (blue line) and ISO (red line) PC1s are shown under the graph.

We have shown that when all spectral features are taken into account, WT and ISO songs are indeed different.

1.4 Comparing ISO and WT songs with respect to duration of acoustic state

As we mentioned before, ISO song syllables and notes are often prolonged and monotonic. To quantify this notion, we estimated correlation time, namely, the interval over which acoustic features remain highly correlated. Song correlation time can be calculated by computing the spectral auto-correlation of the song bout, and measuring the intervals, starting from the diagonal of the auto-correlation matrix, where the correlation coefficient is higher than a certain threshold. However, during harsh and noisy sounds (e.g., a long monotonic buzz), spectral correlation time is short, even though the bird does not change the features of his singing, so the articulatory state probably remains unchanged. Using features that correlate with the articulatory state provides us with better estimates of correlation time. We therefore calculated the duration of acoustic state based on pitch, FM, Wiener entropy & goodness of pitch. We called this feature the duration of acoustic state (DAS), because it is an indication of the time period that a birds spends singing a sound that in relatively unchanging. We developed an algorithm by modifying an earlier algorithm (Tchernichovski et al 2001) used to calculate the period of repetition in songs.

We scaled the features to units of statistical distances (median absolute deviation from the mean, MAD) based on WT zebra finch feature distribution (Tchernichovski et al. 2000). We then constructed a similarity matrix based on Euclidian distances between the spectral frame features, for each spectral frame in the song. For example, starting with a spectral frame t_i, we calculate the Euclidean distance D from the neighboring spectral frame t_{i+1} using Equation 1.

$$D[t_i, t_{i+1}] = \sqrt{\left(pitch_i - pitch_{i+1}\right)^2 + (FM_i - FM_{i+1})^2 + (entropy_i - entropy_{i+1})^2 ...} \qquad \text{Eq. (1)}$$

We then calculate the Euclidean distances from the starting point t_i to its more distant neighbors, namely $D[t_i, t_{i+2}], D[t_i, t_{i+3}]\ldots$, and keep going until the distance crosses a threshold of 2.5 Median Absolute Deviations in 3 consecutive spectral frames. This gives us the right-side duration of acoustic state. We then repeat the same procedure on the left side, namely, $D[t_i, t_{i-1}], D[t_i, t_{i-2}]..$ and combine the left and right side durations to obtain the overall duration of acoustic state (DAS). We consider DAS as undefined when no sound is produced. We implemented this procedure in Sound Analysis Pro 2.

The DAS values for every spectral frame of a zebra finch song motif are shown in Figure 1.10. In this figure, we can see that DAS is an appropriate estimate of the acoustic states of the song, as the values are high for extended periods when the bird produces longer sounds such as harmonic stacks (yellow arrows). In addition, during vibrato sounds, when the frequency modulations are rapid and the momentary spectral features change much, the values largely remain high showing that despite the fluctuations the vibrato state itself lasts long (pink arrow).

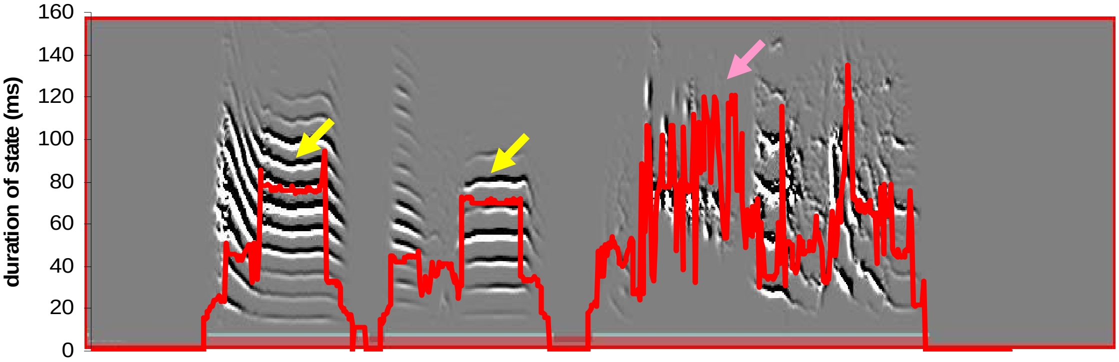


Figure 1.10 | Duration of acoustic state values for a WT zebra finch song. The red line represents the duration of acoustic state values at each time point of the sonogram. Yellow arrows (above first 2 syllables) point at harmonic stacks that have consistently high DAS values. Pink arrow points at a modulated sound (last syllable) where the DAS values remain mostly high indicating that despite the high modulation, the acoustic state in unchanged.

At the beginning of this chapter, we defined a song note as an acoustically uniform unit of song production, so the duration of acoustic state is often similar to note length (Du & Troyer 2006), except that this is a continuous estimate that requires no segmentation of the song.

For every song bout we calculated the values of DAS, which is essentially a long vector just like each of the spectral frame features. We can therefore continue analyzing DAS just as we did with spectral frame features, namely plotting histograms and computing PCA. The probability distribution histograms for the DAS values of ISO and WT birds are shown in Figure 1.11. The histograms show that WT birds (blue line) have shorter durations, in other words, they spend less time in one acoustic state or transition faster between different notes. The ISO curve shows two peaks: the short duration peak is centered on about 40ms, exactly like the WT peak, and the second peak is at about 200ms. The short duration peak is shallower and broader in the ISO songs. The second peak is likely to correspond to the long, unchanging notes often seen in ISO songs (Figure 1.6, Birds 5, 6 & 7). ISO birds may find it more difficult to rapidly switch acoustic states, but whether this is really the case, they show a tendency to get "stuck" singing a particular note.

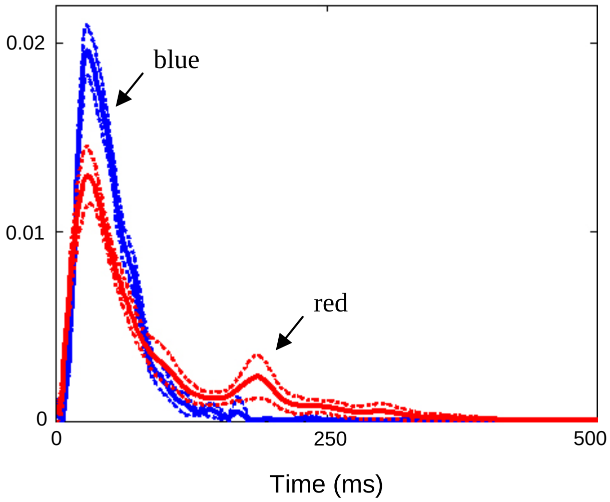


Figure 1.11 | Probability distribution histogram of duration of acoustic state. The red line represents the duration of acoustic state value distributions of ISO birds (n=17), the blue line those of WT birds (n=52). Dashed lines represent 95% confidence intervals.

From the probability distribution histograms, we plotted cumulative histograms (Fig. 1.12) as we had done in our analysis of spectral frame features. For DAS values, the visual advantage of the cumulative histogram over a probability histogram is great as this histogram shows a clear separation between individual ISO birds (red lines) and the confidence interval for the mean of WT birds (dashed blue line). With the exception of perhaps one bird, the red lines (ISO songs) are always below the WT distribution.

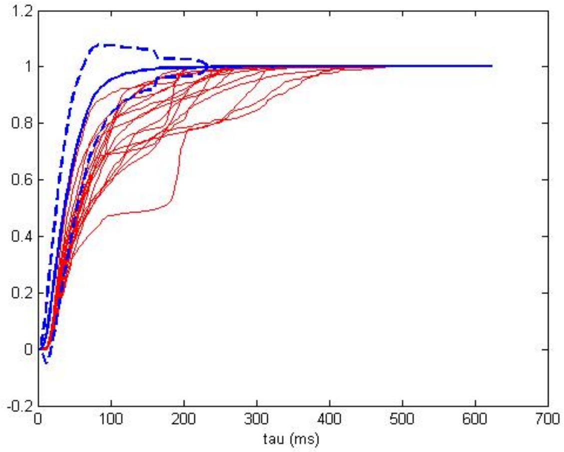


Figure 1.12 | Cumulative Frequency Distributions of acoustic state durations. Each red line represents the song of one ISO bird. The blue line represents the mean values of 52 WT birds, and the dashed line shows 95% confidence intervals.

As with spectral frame features, the DAS histogram is a high-dimensional statistic. To obtain a low dimensional description, we summarized the cumulative DAS distribution vectors using PCA, and plotted the PC1 and PC2 values against each other (Fig. 1.13). Compared to spectral frame features, the separation in the PCA space appears very strong. The inverted U-shape spread of the PCA suggests some correlation between PC1 and PC2, but overall, we see that PC1 separates the WT and ISO distribution much better than PC2, namely all the WT songs are found in the left arm and all the ISO songs are in the right arm. There are only 2-3 birds that are slightly overlapping. The mean values of the first PC were significantly different between ISO and WT clusters ($p<0.001$, t-tests, $n_{WT}=52$, $n_{iso}=17$).

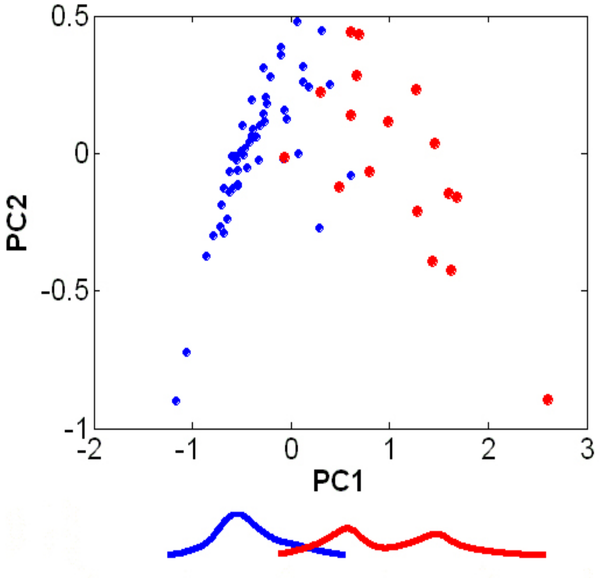


Figure 1.13 | First 2 principal components (PC1 and PC2) of acoustic of state durations of WT and ISO songs. Red (slightly larger) dots represent individual ISO birds (n=17), blue dots WT birds (n=52). Distribution histograms of WT (blue line) and ISO (red line) PC1s are shown under the graph.

1.5 Comparing ISO and WT songs with respect to song rhythm (bout-level song production)

Although both birdsong and human music are characterized by strong rhythms, there are no standard methods for measuring song rhythm. In western music, rhythm is often explicitly annotated, but the question of how to capture rhythm structure when we hear a natural signal that sounds rhythmic remains open. An existing method measures the distribution of intervals between sounds, and looks for regularities in the sequences of intervals (Tierney et al. 2008), but this approach is ad hoc and has limited generality. Recently, Saar & Mitra (2008) proposed a more objective and general method for capturing song rhythm. Their measure aims at describing regularities in the overall bout organization. Here we used this approach to examine differences in rhythm between ISO and WT birds. We have seen in the examples in Section 1.2 that WT songs have stable repeating motifs, but ISO songs often do not. Instead, there is a tendency to repeat syllables, which are sometimes not fully crystallized. Such differences might be captured by the rhythm of songs, and it is perhaps this difference that is most salient when we listen to WT and ISO songs. WT songs sound very periodic and rhythmic, but most ISO songs do not, or they switch their rhythm mid-song.

The Saar & Mitra method is fairly simple: just as frequency analysis of sound waveform in short windows (of say 10ms) can be used to calculate spectral frame features (pitch, FM, etc.), a similar frequency analysis on much longer windows of singing (e.g., an hour) can capture frequency structure in longer timescales. For example, pitch is the estimate of the period of repetition in very short time scales. For example, when we say that harmonic stack in zebra finch song has a pitch of 1000Hz, what we mean is that the waveform within this sound repeats in periods of 1ms (Fig. 1.14).

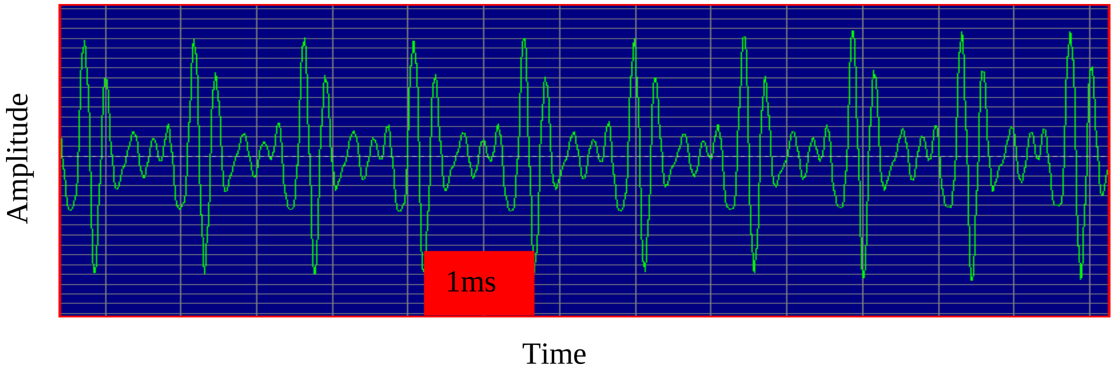


Figure 1.14 | Waveform of periodic sound. Amplitude is plotted against time. Red box indicates the period of the sound.

In a long epoch of singing, when the same motif repeats again and again in a period of 1s, we can also capture the “pitch”, or period of repetition, using frequency analysis in long timescales. Rhythm frequencies can capture patterns of repetitions at both the syllabic level and at the song-motif level. We used a nested spectral analysis method. First, we calculate the spectral frame song features. Then an additional frequency analysis is performed on the song feature time series. The outcome of this second frequency analysis is called *rhythm spectrum*. Although feature values at a given time point depend on the fine temporal structure of the waveform with millisecond resolution, the features themselves change with a slower timescale of 10–100ms. Therefore, the second spectral analysis has most of its energy concentrated on lower frequencies, which allow detection of periodicity in syllable repetition, motif repetition, etc. In the current study, rhythm spectrum was calculated based on the amplitude envelope of song bouts in ISO and WT birds, in the range of 0.5-10Hz.

Figure 1.15 presents the mean and confidence interval for rhythm spectra of WT versus ISO. Note that by averaging spectra across birds we lose most, if not all, information about specific peaks in the rhythm of individual birds, and all we can judge visually is the overall coarse shape of the rhythm spectrum.

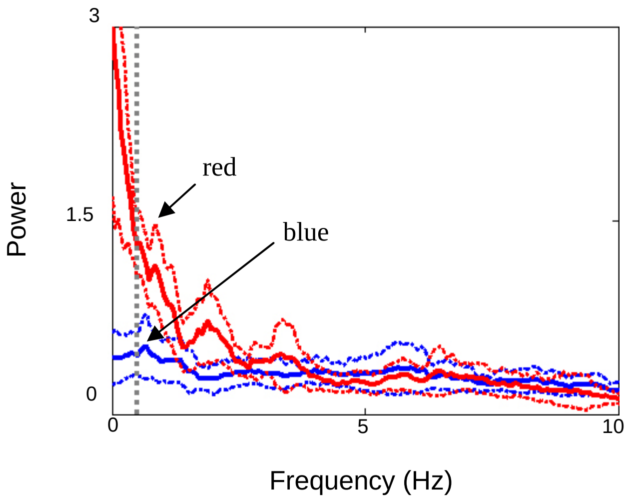


Figure 1.15 | Rhythm spectra for WT and ISO birds. WT (blue, n=52) and ISO (red, n=17) rhythm spectrum distributions. The dashed gray line marks the minimum frequency that we used for further analysis (0.5 Hz).

As shown, ISO songs have much more power in the low frequency range, and the power is going up sharply as we approach zero. There are two possible explanations for this effect: either ISO songs have longer periods of repetitions or the energy we see at the low frequencies of the ISO rhythm spectra is leaking from the zero frequency – which captures the non-periodic energy in the signal. Based on visual inspection of sonograms, which contain lots of shorter rather than longer repetitions (Fig. 1.6 Bird 2) and having listened to many ISO and WT songs, we think that the latter is more likely. Moreover, it could be that ISO song is less rhythmic in the fast repetitions due to jitter in duration of sounds and silence intervals, in which case most of the energy would be concentrated near zero.

As before, we summarize the high dimensional representation (in this case, spectrum) by performing PCA. The results are shown in Figure 1.16. Clearly, the separation in this feature space is not as good as with spectral frame features or duration of acoustic state. Still, the mean values of the first PC were significantly different between ISO and WT clusters for rhythm as well ($p<0.001$, t-tests, $n_{WT}=52$, $n_{iso}=17$).

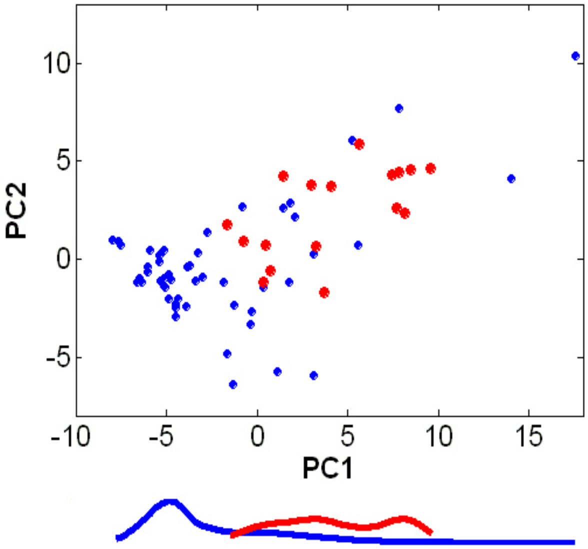


Figure 1.16 | Rhythm spectra PCA for WT and ISO birds. Red dots (slightly larger) represent individual ISO birds (n=17), blue dots WT birds (n=52). Distribution histograms of WT (blue line) and ISO (red line) PC1s are shown under the graph.

In sum, we found significant differences between WT and ISO songs at all timescales of song structure.

1.6 Are the differences due to social isolation or lack of exposure to tutor's song?

Until now, we tested whether ISO songs are different from WT songs. Having confirmed that, we now investigate the cause of this difference. The difference between WT and ISO songs we observed (Figs. 1.5, 1.6) might be caused by the lack of tutoring, but also by social factors unrelated to acoustic experience. For instance, in sedge warblers, deprivation of song tutoring does not result in less structured songs (Leitner et al. 2002). The interpretation of our results would be quite different if the differences between WT and ISO songs using our three song traits (features, duration of acoustic state, rhythm) were mostly due to social isolation, as opposed to the lack of tutoring. To judge this, we used data from birds raised in an arena as shown on Fig. 1.17 (data are from Tchernichovski et al. 2000). We examined if social inhibition of song imitation (Tchernichovski 1998) can lead to ISO-like songs, by allowing 10 pupils to interact with a single tutor simultaneously.

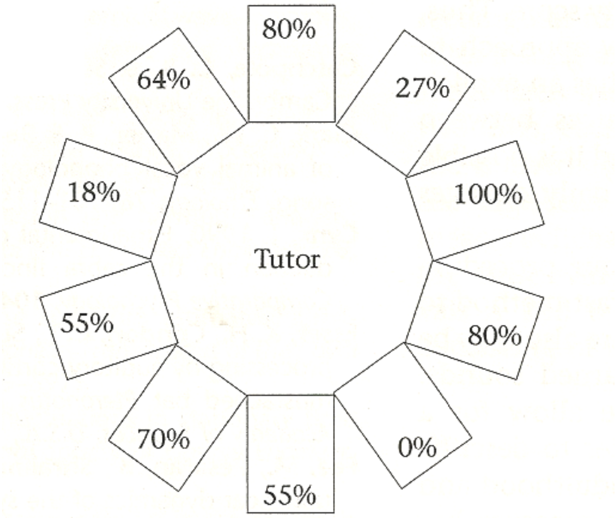


Figure 1.17 | Arena of cages with 10 pupils organized around a single tutor. The tutor was placed at the center, and the 10 pupils around him in a circle. The similarity scores between each pupil and the tutor (manual score) are shown for each pupil.

In such one-to-many arrangements, songs do not converge but diverge. Imitation seems to be inhibited in some pupils but not in others. Given that some birds are more "tutored" than others, we can test if songs that were less influenced by the tutor are more similar to ISO songs. This way, we can test if ISO-like song can emerge in a social situation, therefore excluding the possibility that ISO song is an outcome of social isolation stress. Figure 1.18 presents the similarity to the tutor song against the PC1 of song traits at all three timescales. If lower similarity corresponds to higher PC1 in the ISO range, then we can confirm that ISO song can arise as a result of lower imitation of a tutor and not just in social isolation. The correlation between similarity to tutor's song and the PC1 is statistically significant for spectral features ($p=0.02$) and for duration of acoustic state ($p<0.01$) but not for rhythm. For the first two traits, the PC values of birds that imitated well project on the WT distribution, and the bird that produced the worst imitation projects onto the ISO distribution. For rhythm, however, all birds appear to be within the WT range even though the trend is similar.

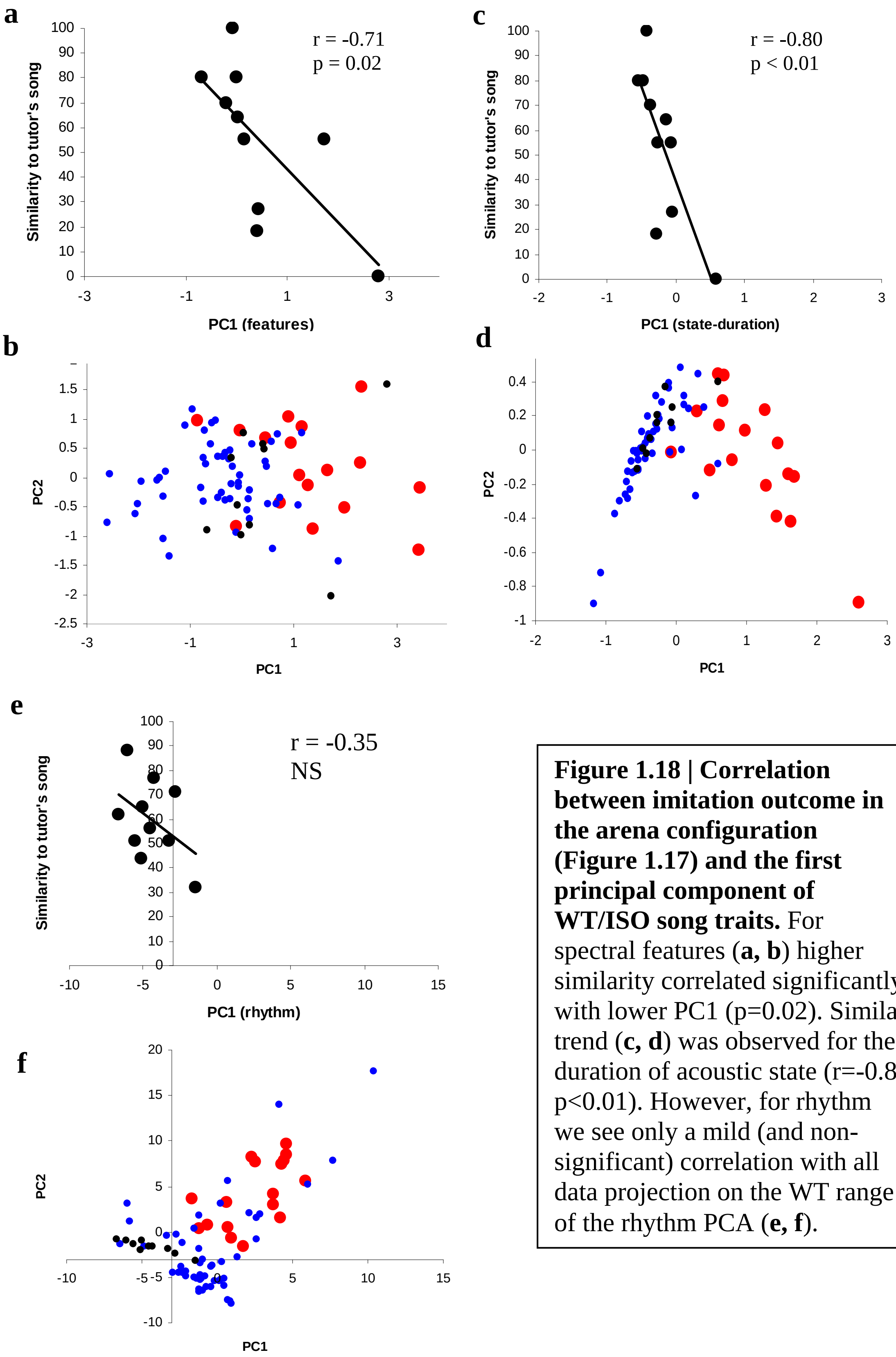


Figure 1.18 | Correlation between imitation outcome in the arena configuration (Figure 1.17) and the first principal component of WT/ISO song traits. For spectral features (**a, b**) higher similarity correlated significantly with lower PC1 (p=0.02). Similar trend (**c, d**) was observed for the duration of acoustic state (r=-0.8, p<0.01). However, for rhythm we see only a mild (and non-significant) correlation with all data projection on the WT range of the rhythm PCA (**e, f**).

In summary, the songs were WT-like in pupils who imitated accurately, but more ISO-like the less the pupil copied from the tutor, confirming that zebra finch ISO songs are, to a large extent, an outcome of tutoring deprivation. There might be an additional effect of isolation stress on zebra finch song development, and such an effect can be tested by raising birds by mute parents.

1.7 Conclusions

In this chapter we showed that when zebra finches are deprived of song tutoring, they improvise a song that is different in many ways from regular WT song. Although there is great diversity in ISO song, deviations from typical WT songs were found at three timescales of song structure: in ms-to-ms spectral features, in the duration of acoustic state and in song rhythm. Differences were significant at all timescales. We will now examine the songs of birds trained by isolates and will use the measures we introduced in this chapter to test if those songs are similar to ISO or WT songs.

Chapter 2. Imitation of ISO song

Background & Rationale

Songbirds are masters of vocal imitation, capable of copying adult song with a nearly perfect accuracy. However, recent studies showed that despite the capability to imitate accurately, birds sometimes deviate from their model song. Examples include the social inhibition of imitation in zebra finches (Tchernichovski 1998) discussed in Chapter 1. Another revealing experiment in canaries took advantage of the fact that canary song is organized into phrases (repetitions of identical syllables) that make up the song. This syntax structure is found even in isolates. Gardner et al. (2005) trained young canaries with synthetic songs that consisted of syllable shapes comparable to normal canary song but organized in a "random walk" fashion. So rather than identical syllables and abrupt phrase boundaries, the syllables gradually turned into another type. The young birds learned these songs but as they matured, they switched to the species-typical phrase structure. Therefore, imitation and innate constraints are separate processes and can be segregated in time. In young birds, imitation prevails, but later in life, the innate biases come to dominate.

White-crowned sparrows are able to assemble complete songs when tutored only with phrase pairs (Rose et al. 2004). Although there is a strong innate predisposition to start a song with a whistle, young birds can be trained to sing in reverse order (whistle coming last) if trained with phrase pairs ordered in such a way. However, when trained with overlapping phrase pairs in the natural, forward order (AB, BC, CD, DE), birds easily assembled the whole sequence together (ABCDE). Birds trained with single phrases did not produce complete forward-ordered songs, although in 8 out of 9 cases they started their songs with a whistle (A) followed by a B-type phrase, which probably shows the extent of their innate predispositions. This example shows that imitation is aided by innate predispositions but can override them in case of a conflict.

There are two opposing forces acting on song learners that can have profound effects on the evolution of song cultures. On the one hand, random inaccuracies (“song-mutations”) and innovation among the young learners are the source of diversity, providing the raw material for the formation of local dialects. They can also facilitate the formation of individually distinct songs (perhaps indicating individual identity) within the dialect. On the other hand, biases in imitation (as in Gardner et al. 2005) might mirror innate constraints that keep song dialects within the WT range, otherwise imitation errors and innovation would cause unbounded variation over many generations or large geographical distances, which does not seem to the case (Marler & Tamura 1962). These postulated imitation biases could either alter the imitation of song features in a certain direction, or make it more likely that young birds imitate certain parts or certain features of songs.

What happens if we provide a young bird with an ISO tutor? Will the young bird exhibit imitation biases by preferential imitation of WT-like syllables or WT-like song features? We already know that juvenile males accept ISO song as a valid song model, and they imitate them even in the presence of WT tutors (Williams 1993). Moreover, in a colony, sons of ISO fathers imitate more of their father’s songs than sons of WT birds do of their own father’s songs (Williams 1993). However, studies that investigated the imitation of ISO song used coarse analysis methods, such as counting syllables in the pupil’s song that appear to be similar to those in the tutor’s song, and small or moderate imitation biases might not be captured using such methods.

In this dissertation, we tried using fine-grained quantitative methods, like the one we developed in Chapter 1 for assessing distances between WT and ISO songs. In Chapter 1 we showed that WT and ISO songs form distinct clusters along a continuum at the three timescales of song structure we investigated: spectral features, duration of acoustic state and song rhythm. We now examine the imitation of ISO songs and attempt to answer the question whether pupils of isolates show biased imitation regardless of tutor song and individual identity, and if they do this by selective imitation of WT-like syllables or modifications to existing syllables.

Methods

Animals

We trained 13 juvenile birds by ISO tutors one-to-one. We randomly selected hatchlings from 40 breeding pairs and did not let them hear songs after day 7 post-hatch. At this time, they were separated from their fathers and moved to a separate nursery area with their mothers and siblings. On day 30, the mothers were returned to the fathers and the juvenile males were paired with one of 6 isolate tutors. These tutors were randomly selected from our 17 ISO birds (isolation procedures are described in Chapter 1). The ISO tutor and his pupil were kept together for 90 days in a sound isolated chamber. To evaluate the effect of individual tutors, four of the isolate tutors were used 2-4 times to train unrelated pupils. Tutors' ages ranged between 140-1571 days (median age at beginning of first tutoring = 316 days). Additional training was performed serially (one-to-one) after confirming that the isolate tutor song remained stable over the tutoring period based on visual inspection & feature distributions.

Recording and Analysis

Sound recording was continuous during training using Sound Analysis Pro. However, since there were two sources of songs in every chamber (the pupil and the tutor), to obtain single recordings from every bird for positive song identification, we placed the pupil in a separate sound chamber, subsequent to the maturation of every pupil (around day 120), and recorded his song. Further analysis was done on these single recordings.

Data analysis & statistics were done using Matlab 7.0. Calculations of spectral features, duration of acoustic state and similarity measurements were done by Sound Analysis Pro 2. Detailed statistical tests are included in Appendix III.

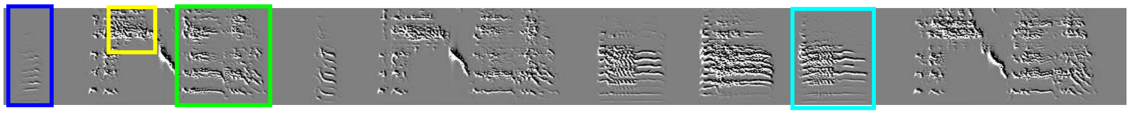
Results

2.1 Visual assessment of imitation

First, let us examine the outcome of ISO tutoring by visually inspecting the sonograms and comparing tutor songs to all of their pupils' songs.

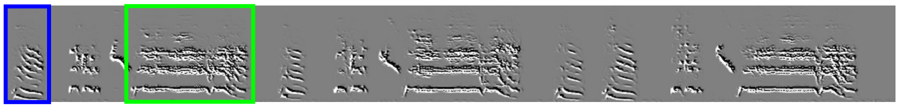
ISO Tutor 1 (Bird 19)

First generation Pupil 1 (Bird 1248)



First generation Pupil 2 (Bird 1302)

First generation Pupil 3 (Bird 1340)



First generation Pupil 4 (Bird 1661)

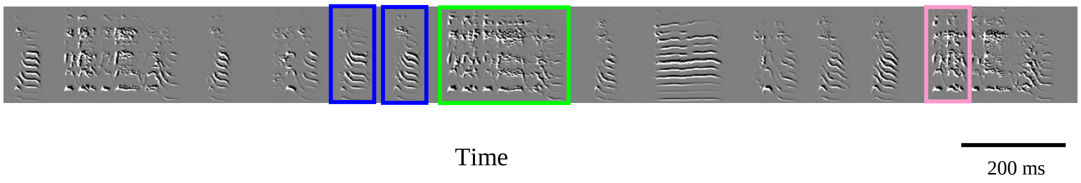


Figure 2.2 | Sonograms of Tutor 1 and his pupils. Song notes are marked with different colored rectangles (blue, pink, yellow and green in order in ISO song).

Tutor 1 (Fig. 2.2 upper panel) was our oldest tutor bird (age=1571 days at start of first tutoring) and his song was highly stable. He had a one-syllable motif and a rather stable bout structure. In addition to the simple bout structure, some of the notes are clearly ISO-like, with a narrowband noisy note (in yellow) followed by a short, high-pitched note and then a broadband noisy harmonic note (which sounds like a scratch). As seen above, all of the pupils of Tutor 1 imitated his song with more or less accuracy. Pupil 4 shows the least similarity. In his song, the only recognizable elements are the introductory syllable (blue rectangle) and the last, messy harmonic (green rectangle), but it seems like the pupil constructed this syllable from the spectral material of the first note of the complex syllable (pink rectangle). The duration of the yellow note, as well as that of the green one, is decreased in the songs of Pupil 1 & 2. Pupil 3 omitted the yellow note altogether. The syllables that dominate the song of the ISO tutor tend to be less abundant in his pupils – who added introductory notes and calls to the motifs. These songs and calls can all be found in the tutor's repertoire, but the tutor, in contrast to the pupils, sings them rarely.

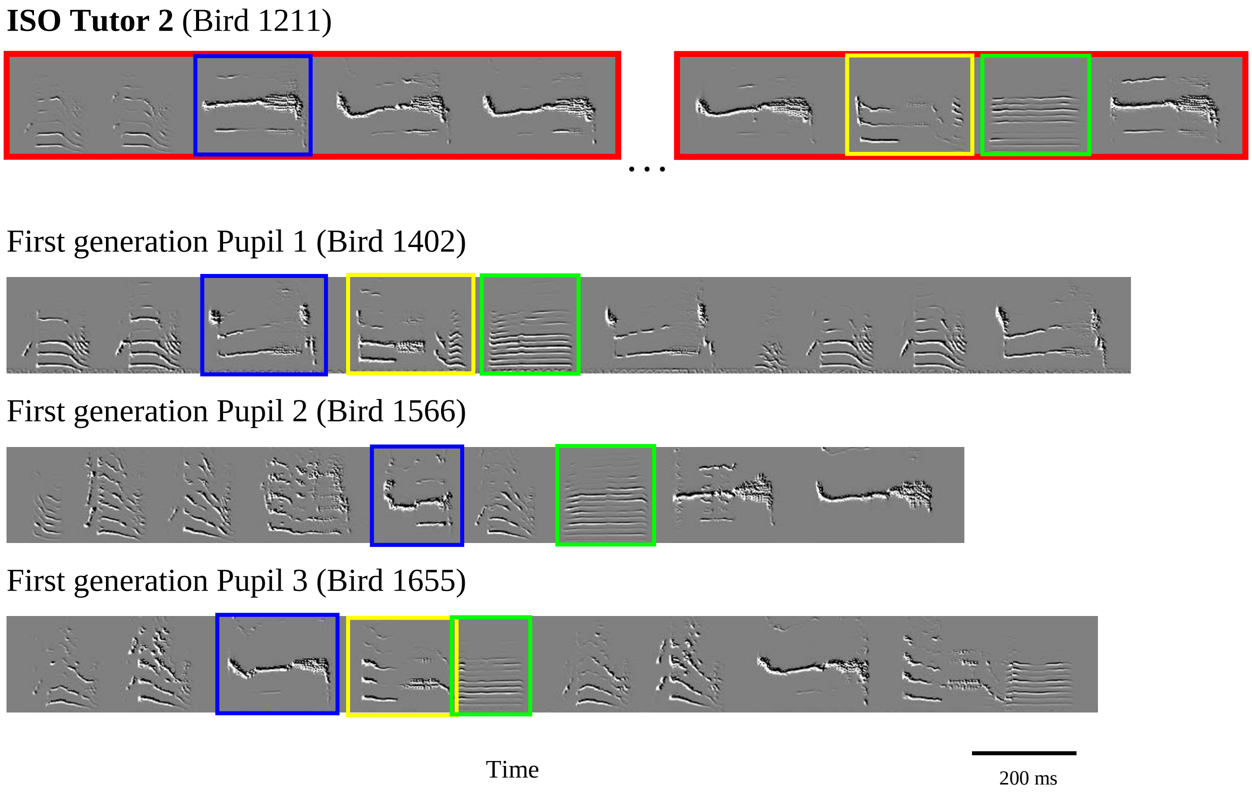


Figure 2.3 | Sonograms of Tutor 2 and his pupils. Important syllables are marked with different colored rectangles (blue, yellow and green in order).

Tutor 2's song (Fig. 2.3) contains one long (230 ms), high-pitched note (blue) that is very atypical of WT zebra finch song and two other syllables of similar duration that are harmonic-like (yellow and green). Pupils of Tutor 2 also imitated most of his song syllables. What is striking here is that none of the pupils produced repetitions of the blue syllable even though the tutor repeated this syllable at least 4 and up to 24 times in a song bout (median repetition rate=15). On the other hand, the syllables that were rare in the tutor's song (yellow and green squares) were much more frequent in all of the pupils' imitations. Pupil 2 shortened the blue syllable and Pupil 3 fused the yellow and green syllables.

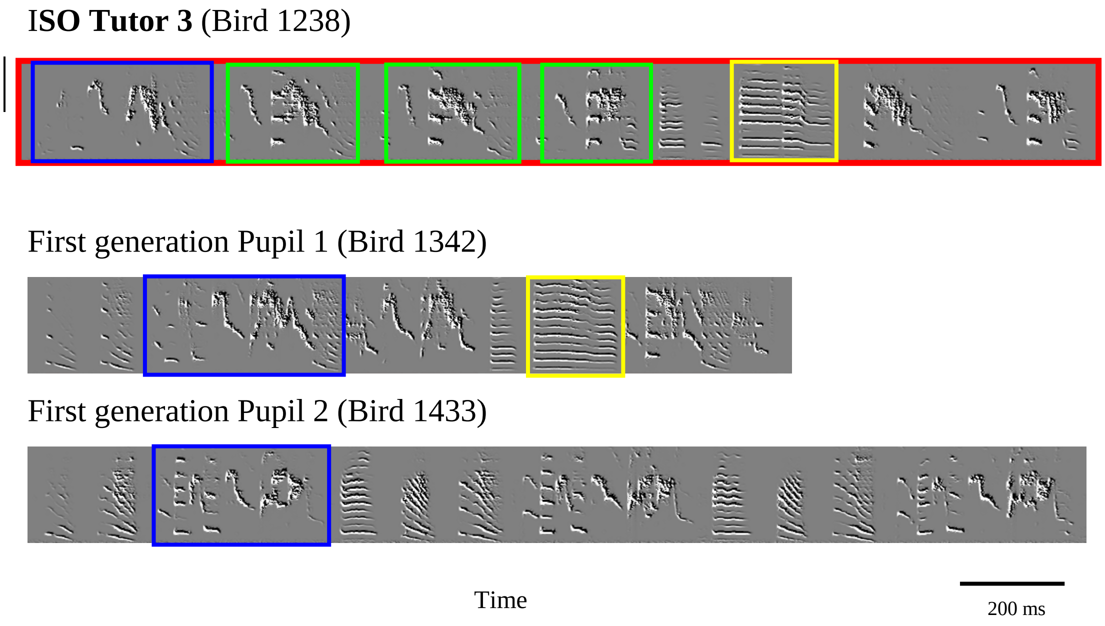
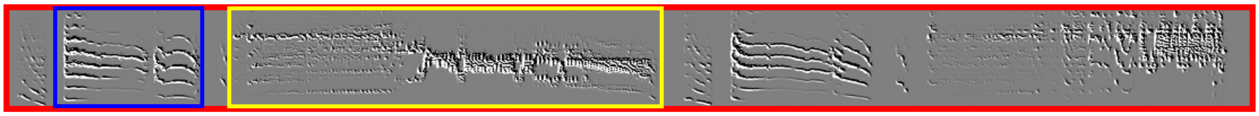


Figure 2.4 | Sonograms of Tutor 3 and his pupils. Important syllables are marked with different colored rectangles (blue, green and yellow in order).

Tutor 3 produced a song that is quite complex compared to other ISO songs (Fig. 2.4). For instance, it has fast transitions between different notes whose spectral shapes are quite complex. However, it does contain some features that are typical only to ISO song, such as the frequent back-to-back repetition of one syllable (green rectangles). Also, the harmonic sequences of syllables that appear occasionally in the bout are atypical e.g., the last harmonic syllable is long and includes a short stop in the middle, followed by a low amplitude ending (yellow). Both of these ISO-like features were changed by the pupils. We can observe reduction in Pupil 1 and omission in Pupil 2 of the long harmonic syllable, and a decrease in the number of the complex syllable repetitions to 2 Pupil 1 and 1 in Pupil 2.

ISO Tutor 4 (Bird 1247)



First generation Pupil 1 (Bird 1315)

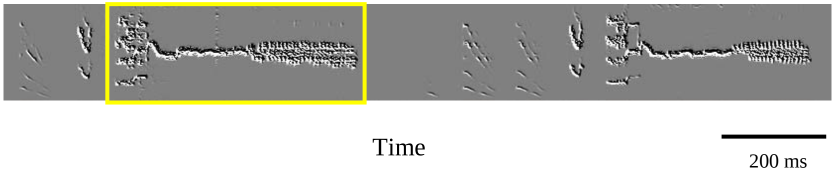
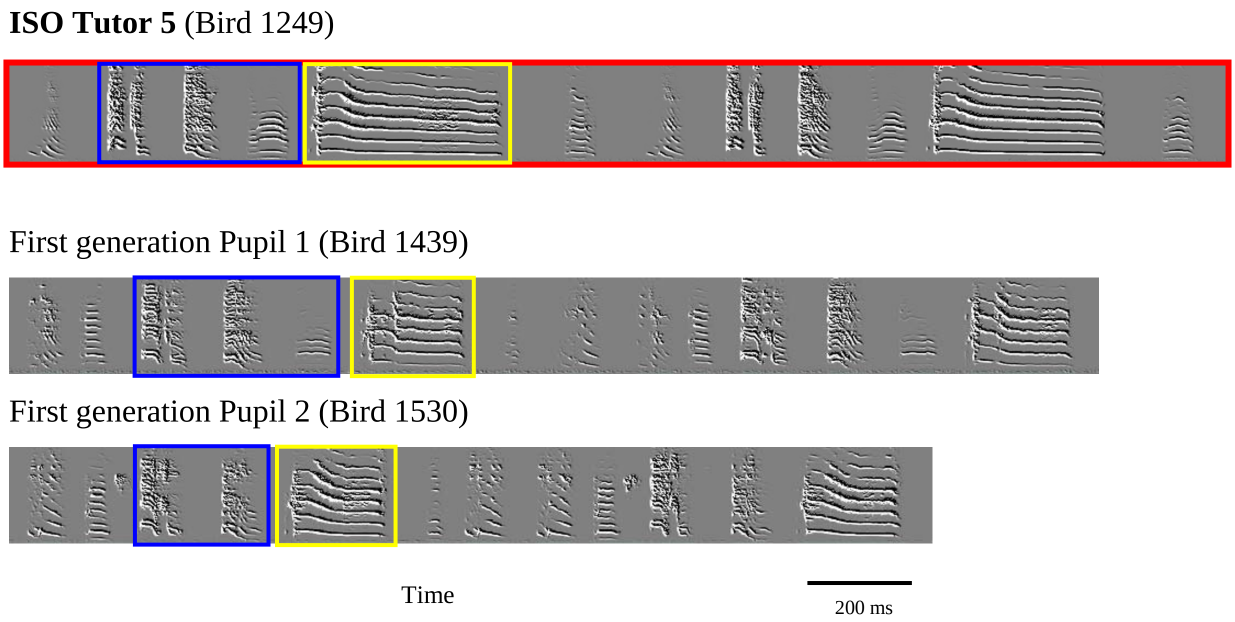


Figure 2.5 | Sonograms of Tutor 4 and his pupils. Important syllables are marked with different colored rectangles (blue and yellow in order).

Tutor 4 had the most abnormal song among our tutors (Fig 2.5). There is a harmonic element (blue square) that is long and call-like (although it was not included in every single bout) followed by an unstructured, broadband, scratchy syllable (in yellow) that is more than a second in duration, which is longer than the typical zebra finch song motif. This syllable is highly variable in the tutor in both duration and bandwidth and internal structure. The pupil of this tutor did not imitate the call-like harmonic, and greatly reduced the length and the bandwidth of the scratchy syllable. He has created a stable motif and bout structure out of a highly unstable song.



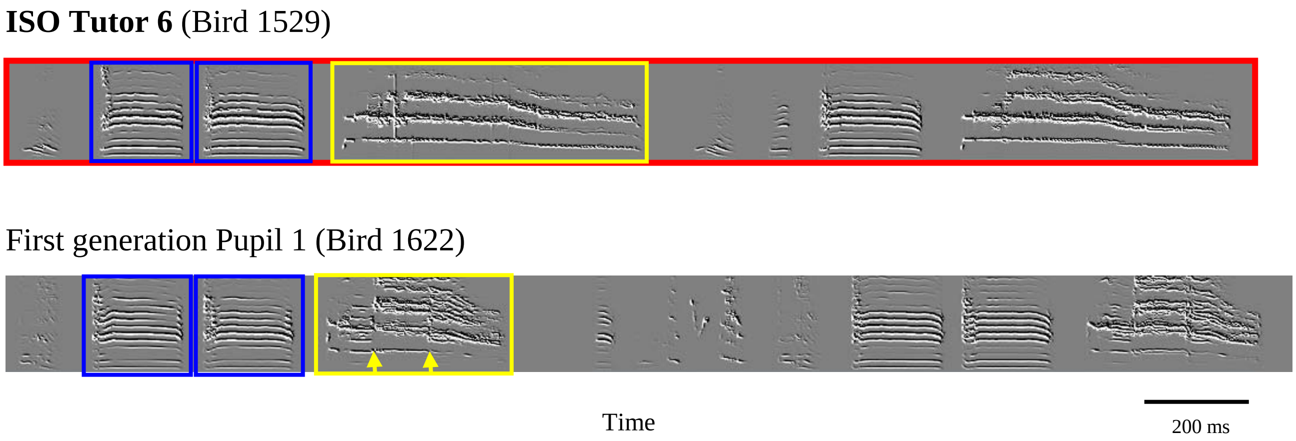
ISO Tutor 5 (Bird 1249)

First generation Pupil 1 (Bird 1439)

First generation Pupil 2 (Bird 1530)

Figure 2.6 | Sonograms of Tutor 5 and his pupils. Important syllables are marked with different colored rectangles (blue and yellow in order).

Tutor 5 has a song (Fig. 2.6) with a few short syllables (blue rectangle) that are within the normal, WT range in duration as well as other spectral features, but the last syllable (in yellow) is an extremely long, call-like harmonic. Both pupils of this bird shortened the long call into a medium-length harmonic (a typical element of WT song) but copied all or most of the other syllables with high accuracy.



ISO Tutor 6 (Bird 1529)

First generation Pupil 1 (Bird 1622)

Figure 2.7 | Sonograms of Tutor 6 and his pupils. Important syllables are marked with different colored rectangles (blue and yellow in order).

Tutor 6 produced a very simple and highly abnormal, though quite stable, song. There are two long harmonics (blue) and an extremely long, high-pitched harmonic (yellow). The pupil imitated the harmonics accurately but shortened the long syllable. He not only shortened it, but also started to differentiate it into notes (boundaries marked with yellow arrows).

Summary of subjective inspection

Overall, there was a general tendency among pupils of ISO birds to shorten certain syllables that were of long duration. Repetitions of syllables disappeared in the songs of the pupils and some syllables became more complex or more stereotyped. In all cases, the pupils' motifs were highly stable: they always repeated the syllables in the same sequence with only minor variations, even in cases where their tutors sang extremely variable songs. Scratchy sounds were not imitated or greatly reduced in bandwidth and harmonic structure tended to be more stable in pupils than ISO tutors.

Before a formal test of whether these deviations from ISO tutor songs were robust and significant, we present a numerical summary of the imitation accuracy and syllable identities in tutor and pupil songs.

2.2 Imitation statistics

We used Sound Analysis Pro 2 to assess the accuracy of imitation. We segmented tutor and pupil songs into syllables and calculated imitation accuracy by using symmetric comparisons and then averaging across syllables. The determination of syllable types and their equivalents in pupils was done subjectively, but we confirmed our choices with accuracy measurements between all tutor and pupil syllables. Then we categorized syllables into complex or call-like syllables, once again subjectively, and calculated imitation rates for the different types of syllables. Table 2.1 contains the imitation statistics for each tutor and pupil pair.

Tutor	**19**	**19**	**19**	**19**	**1211**	**1211**
Pupil	**1248**	**1340**	**1302**	**1661**	**1402**	**1566**
Tutor: complex	1	1	1	1	3	3
Pupil: complex	1	1	1	1	3	3
% copied	100	100	100	50	100	100
% invented	0	0	0	0	0	0
Accuracy	81	63	72	39	88	76
Tutor: call-syll	0	0	0	0	0	0
Pupil: call-syll	1	1	1	3	2	3
Tutor: rare syll	2	2	2	2	2	2
#calls copied	3	?	?	?	2	2
#callsinvented	0	0	1	1	0	1

Tutor	**1211**	**1238**	**1238**	**1247**	**1249**	**1249**	**1529**
Pupil	**1655**	**1342**	**1433**	**1315**	**1439**	**1530**	**1622**
Tutor: complex	3	4	4	1	2	2	1
Pupil: complex	3	4	3	1	2	2	1
% copied	100	100	75	75	100	100	100
% invented	0	0	0	0	0	0	0
Accuracy	85.5	85.5	81	49	90	75	75.5
Tutor: call-syll	0	3	3	0	3	3	3
Pupil: call-syll	0	2	3	1	3	2	2
Tutor: rare syll	2	0	0	2	0	0	0
#calls copied	2	0	0	1	3	2	2
#calls invented	0	0	0	0	0	1	1

Table 2.1 | Imitation of isolate tutors in each tutor/pupil pair. Columns show the individual statistics for every tutor and his pupil. Per cent copied is the percentage of ISO tutor syllable types that were clearly copied by their pupil (judged subjectively), and where similarity measurements with SAP showed at least 50% significant similarity across the two syllables.

Table 2.1 shows that, with a couple of exceptions, imitation level was quite high, with moderate accuracy. Innovation was low and complex syllables were imitated in almost all cases. More birds invented call-like syllables than complex syllables.

Next, we examine these statistics across all tutor/pupil pairs. We calculated averages of imitation accuracy and imitations of complex and call-like syllables, and we summarized the results in Table 2.2.

	average	**CV**
Tutor: number of complex syllable types	1.85	0.58
Pupil: number of complex syllable types	1.92	0.54
% of complex syllables copied	92.31	0.17
% of complex syllables invented	5.08	2.44
Accuracy (mean across all syllables)	72.73	0.20
Tutor: number of short/long call types within motif	1.15	1.32
Pupil: number of short/long call types within motif	1.85	0.53
Tutor: # of rare syllables	1.23	0.82
# of calls + rare syllables copied	1.70	0.62
# of call types invented by pupil	0.38	1.32

Table 2.2 | Imitation of isolate tutors across all tutor/pupil pairs. Columns show the averages across all birds and the coefficient of variation. In the ISO tutors, “within motif” refers to syllables that appear in the middle of singing bouts (without an attempt to define motifs).

Table 2.2 shows that pupils imitated nearly all complex syllables (>92%) and their songs contained more complex syllables and less call-like syllables than the songs of their ISO tutors. Overall imitation accuracy was quite high (>72%), and there was little invention, but more complex than call-like syllables were invented.

2.3 Approximation of WT song features in pupils of isolates

In Chapter 1 we presented our results using PCA. We continue using PC1 and PC2 of song features at all three timescales to test whether pupils’ imitations of ISO song significantly differ from our baseline of ISO songs (Fig. 2.8). PCA reveals that pupils are indeed intermediate between WT and ISO songs. The mean values of PC1 for the first generation pupils differed significantly from both ISO and WT means for the spectral-frame features and for DAS (p=0.018-0.001, n=13), but not for rhythm. Feature distributions of most individual pupil songs were closer to WT songs than were their tutor’s songs (12/13 at at least one timescale, 10/13 at all timescales, FDR significance=0.01, binomial test, n=52, Appendix III).

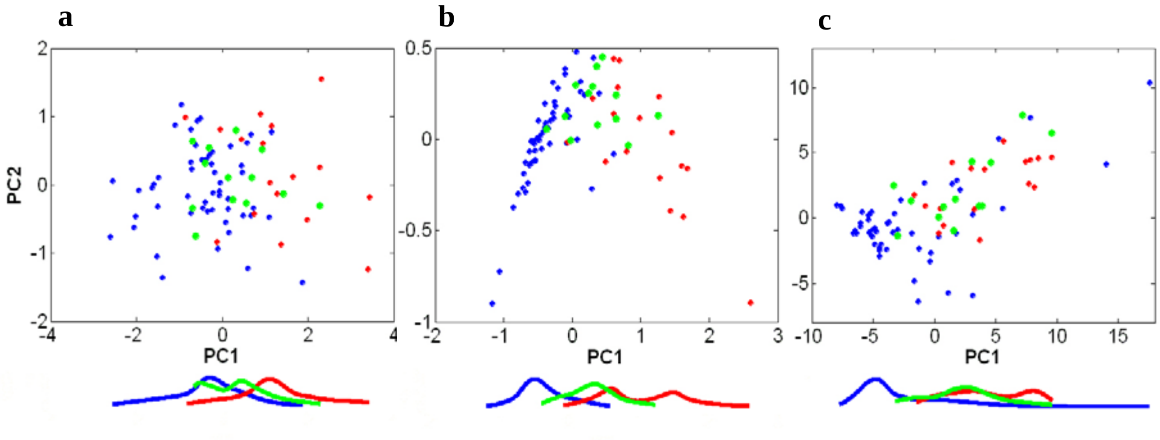


Figure 2.8 | Pupils imitating ISO song approximate WT distribution. PC1 and PC2 of combined spectral features (**a**), duration of acoustic state (**b**) and rhythm (**c**) in ISO birds (red dots, n=17), WT birds (blue dots, n=52) and pupils of isolates (green dots, n=13). Green dots fall between blue and red dots.

After confirming that imitation of isolate tutors is robust and significant, we now turn to examine the details and the possible biases in the imitations, using the methods developed in Chapter 1. We investigate the pupils’ imitations at all three different timescales of song structure: spectral features, duration of acoustic state and song rhythm.

2.4 Spectral frame features in pupil songs

Spectral frame features describe the moment-to-moment changes in song. The three spectral features that provided the best separation between ISO and WT song were amplitude modulation (AM), frequency modulation (FM) and goodness of pitch. We will use these three features to investigate imitation of ISO song. Figure 2.9 presents the probability distribution histograms (top row) and the cumulative distribution histograms (bottom row) for all ISO birds (red lines), WT birds (blue lines) and pupils of isolates (green lines). These plots show that ISO sounds are lower in AM in general (Fig. 2.9a,b higher peak around zero in ISO), have more low-FM sounds (Fig. 2.9c,d high peak near zero in ISO) and lower goodness of pitch (high ISO peak near zero to the left of the WT peak Fig. 2.9e,f). In all panels, the green curves (pupils of isolates) appear to lie between red (ISO) and blue (WT) curves.

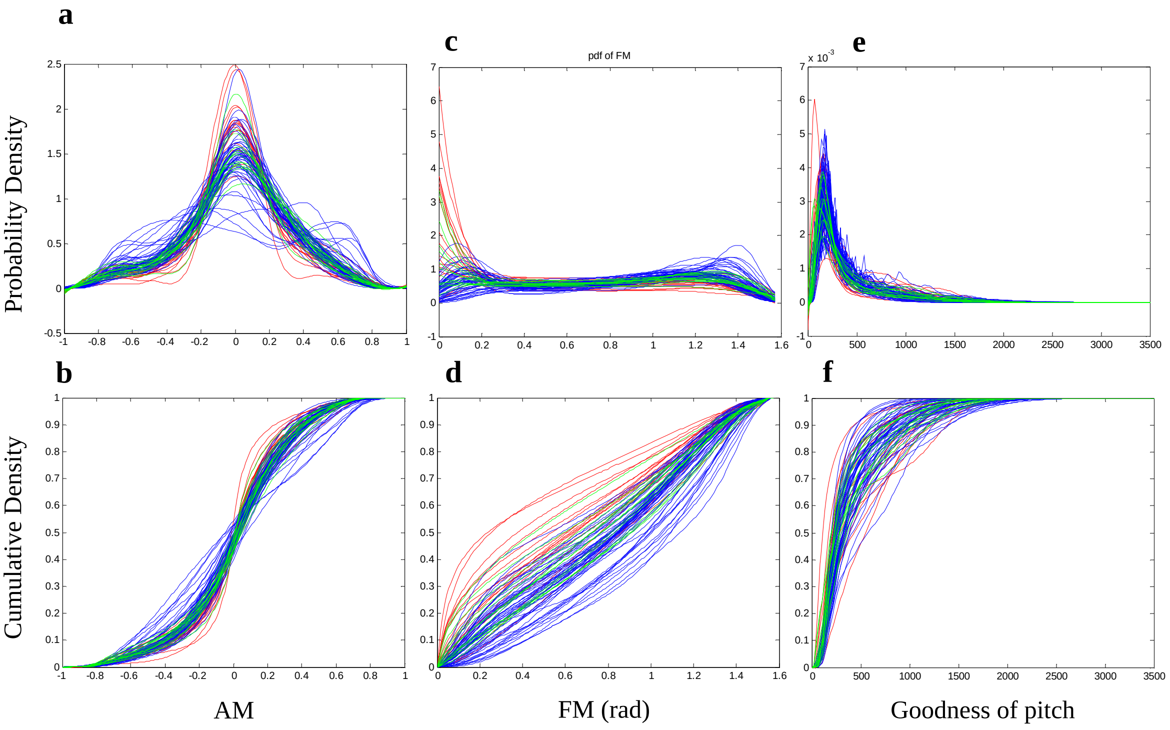


Figure 2.9 | Probability and cumulative distribution histograms of three spectral features in pupils of ISO tutors: AM (a,b), FM (c,d) and goodness of pitch (e,f). Red lines represent ISO birds (n=17), blue lines WT birds (n=52), green lines pupils of isolates (n=13). Probability distribution histograms in top row, cumulative histograms on bottom. (See Chapter 1 Methods for information on units of AM and goodness of pitch.)

In our further analysis, we use the scaled combination of the three features presented in Fig. 2.9. We used the PCA coefficients established for WT and ISO birds to project the song feature distributions of the first generation pupils on the PCA space of the WT/ISO songs (Fig. 2.10). The red dots represent isolate birds, the blue dots wild-type birds, and the green dots pupils of isolate tutors. The green dots appear intermediate between the red and the blue clusters, indicating a shift toward wild-type spectral frame features in pupils of isolate tutors. Histograms of PC1 distribution for each group (ISO, WT & pupils) are shown below the PCA plot, confirming that spectral feature distributions moved toward WT at all timescales. The mean values of PC1 for the first generation pupils differed significantly from both ISO and WT means for the spectral-frame features (p values 0.018-0.0126, respectively, n=13 pupils).

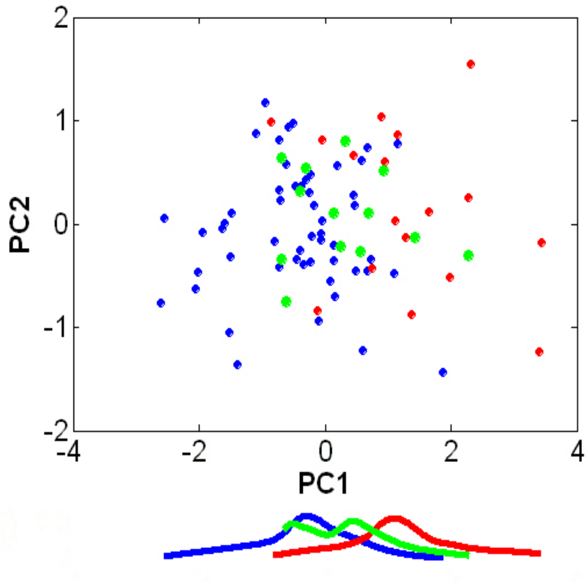


Figure 2.10 | PCA distribution of pupils of isolates is intermediate between ISO and WT distributions. PC1 and PC2 of spectral features of ISO (red, n=17), WT (blue, n=52) and first generation pupils of ISO birds (green, n=13). Histogram of PC1 values at the bottom. Shift towards WT features is indicated by the green dots positioned between the red and the blue dots.

Now we have seen that, overall, the spectral features of pupils are closer to WT spectral features than those of their ISO tutors. However, there has been no information so far about the effect of individual tutors. To assess whether there was progression towards WT in each case of tutoring, we drew arrows from each ISO tutor to all of his pupils (Fig. 2.11). Pupils of the same tutor are labeled by the same color. The location of the WT cluster in the PCA space is indicated by purple shading (shading represents the center of WT distribution based on density estimates; faint dots represent individual WT birds as in Fig. 2.10). As shown, most arrows point in the direction of the WT cluster. We did not observe any idiosyncratic effects of individual tutors on the size or direction of arrows, although pupils of ISO birds who were farther from the WT distribution made a longer shift in general (blue arrows) and variability across pupils (judged by the angle between arrows) was moderate. The only pupil that did not show a clear shift towards WT spectral features was the pupil of Tutor 4 (green arrow). This pupil imitated an extremely abnormal song (Fig. 2.5), and although his copy of the tutor syllable was shorter and of reduced bandwidth, the high-pitched extended note he produced is still quite abnormal. In some cases, like in the pupils of Tutor 2 (blue arrows), the shift towards WT-like features was very large. This was the tutor who sang the same high-pitched whistle-like syllable over and over in his song (Fig. 2.3), but his pupils reduced the pitch and emphasized his more WT-like syllables (yellow and green rectangles in Fig. 2.3) in their songs.

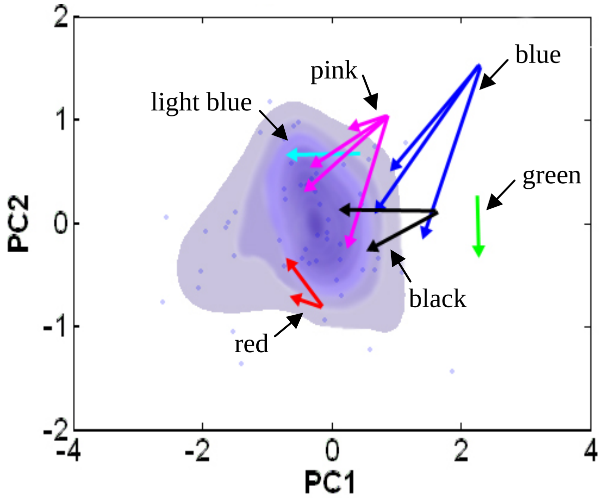


Figure 2.11 | Pupils of isolates shift towards the WT distribution. Different arrows indicate different tutor/pupil pairs, different colors different tutors. Blue shading shows the density estimates of the center of WT distribution.

We have seen that pupils showed a tendency to omit or alter ISO notes of certain spectral shape – e.g., “buzz” notes (ISO Tutors 1 & 4). Indeed, spectral features as judged by the first PC of the feature distribution were not copied accurately. Linear regression analysis of pupil versus tutor values in the first generation yielded a nonzero intercept and a slope slightly less than one (Fig. 2.12). The equality line, corresponding to faithful copying (pupil=tutor, dashed blue line), was rejected in favor of the alternative hypothesis represented by the linear fit shown in red ($P<0.001$, likelihood ratio test, $n=13$). Although copying is not accurate, it is consistent across tutors and pupils (as evidenced by the liner fit), so we can conclude that the imitation errors reveal universal biases.

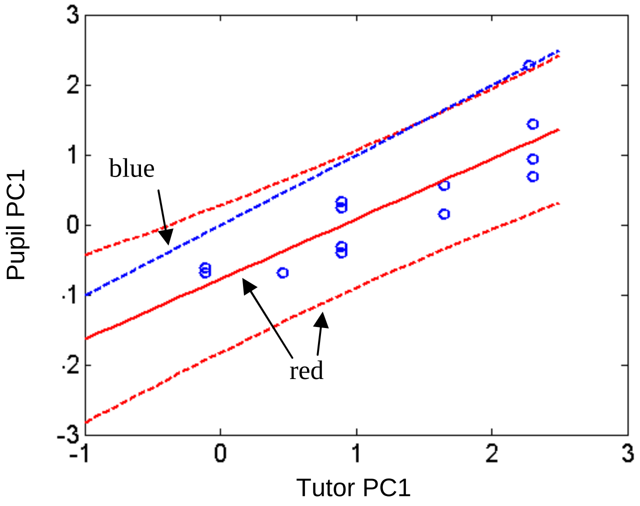


Figure 2.12 | Biased copying of spectral features. Correlation between first PCs of pupil versus tutor, indicating biased imitation. Dashed red line represents 95% confidence band, and the dashed blue line is the identity line.

2.5 Syllable and note duration in pupils of ISO birds

In this section, we will assess imitation in pupils of ISO birds with respect to deviations in note and syllable length as compared with their tutors' songs. We will investigate this question three different ways: by measuring individual syllable durations in tutors' and pupils' songs and comparing them, by calculating durations of acoustic state and by comparing longest to shortest note durations in ISO tutors and their pupils.

A. Syllable duration

When inspecting the sonograms, we noticed a common trend among pupils of ISO tutors, which was the decrease in the durations of abnormally long syllables (Figs 2.2, 2.5, 2.6, 2.7). To test for such an effect quantitatively we first identified syllables in tutor and pupil songs (as in Table 2.1) and then for each imitated syllable, we compared the duration to that of the tutor syllable. As an example, we present Tutor 5 and one of his pupil in Figure 2.13. Here, the long ISO syllable (red bar, mean duration=367ms, s.d.=29ms) was copied by a pupil, but its duration in the pupil's song was about 30% shorter (mean=243ms, s.d.=7.6ms).

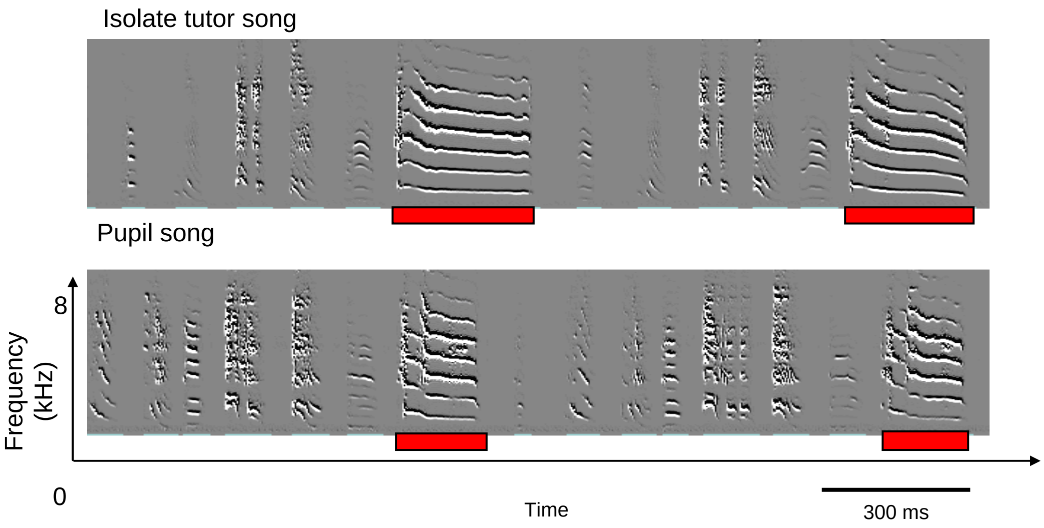


Figure 2.13 | Syllable shortening by pupil of isolate. The abnormally long harmonic syllable (underlined in red) is shortened by the pupil, even though the overall temporal frame (no decrease in motif length) and the durations of other syllables are preserved.

When we examined the relationship between tutor-pupil durations for all the syllables copied across birds, it became clear that not all tutored syllable copies are shorter, and we were interested in finding out whether it is true that the syllable decrease only applies to abnormally long syllables and not to all or any song syllables. Therefore, we plotted the syllable durations of ISO syllables against the durations of the respective imitations of these syllables (red dots in Fig. 2.14). Interestingly, the durations of pupil syllables accurately matched that of the corresponding ISO tutor syllables for syllable durations less than 230ms (r^2=0.98, slope=0.97, n=20 syllables). Copies of longer ISO syllables, however, were invariably shorter than the originals (r^2=0.84, slope=0.56, n=11 syllables). Overall, the range where durations of ISO syllables were accurately copied is similar to the range of WT syllable durations (25-75 percentile range = 67-180ms, n=52 WT birds). It is worth noting that across the entire range of ISO syllable durations, tutor-pupil durations were linearly related on a logarithmic scale (r^2=0.95, slope=0.84, n=31 syllables).

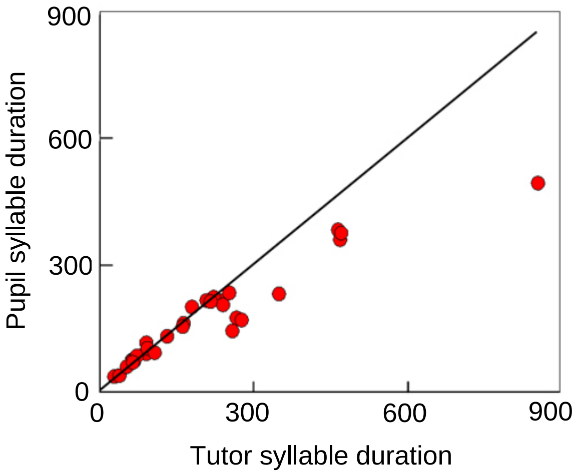


Figure 2.14 | ISO tutor syllable duration vs. duration of same syllable in pupil's song. Pupils faithfully copy syllable durations as long as they are less than about 230 ms. Longer syllables become shorter in pupils' songs.

B. Duration of acoustic state

To judge more generally if ISO songs change their acoustic states more slowly compared to WT birds, we developed the durations of acoustic state (DAS) measure (Chapter 1). We now examine if songs of pupil of ISO tutors are more similar to WT songs in DAS. Figure 2.15 presents cumulative distribution histograms ISO and WT birds and pupils of isolates. The green lines in Fig. 2.15b, which represent the pupils of isolates, fall in between the red lines (ISO birds) and the blue lines (WT birds). Therefore, we can conclude that there is a shift towards WT at the level of the note, too.

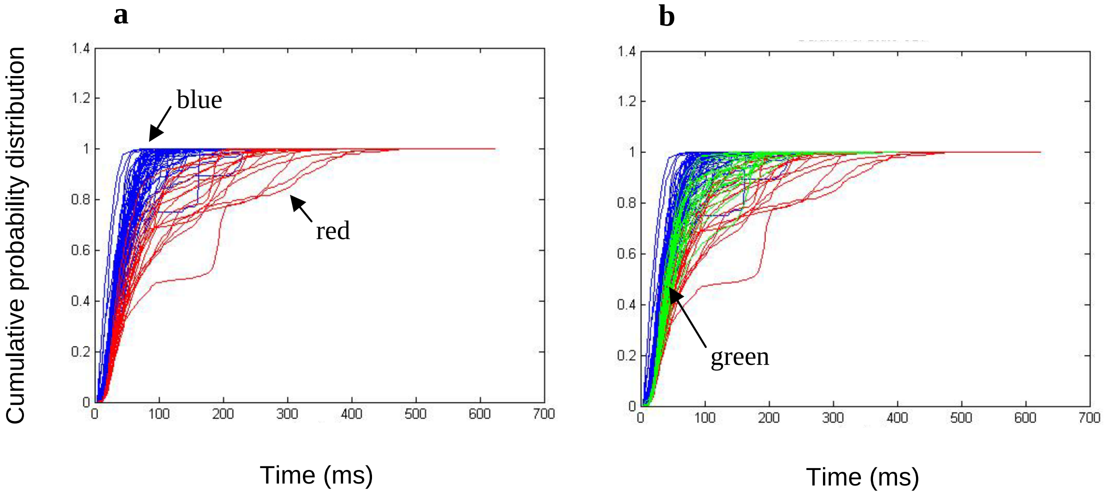


Figure 2.15 | Cumulative probability distributions of acoustic state durations in ISO birds, WT birds and pupils of isolates. a. Each line represents an individual ISO (red, n=17) or WT (blue, n=52) bird. **b.** Pupils of isolates are added in green (n=13).

Once again, we performed PCA to provide a better visualization (Fig 2.16) and statistical analysis. Results confirmed a significant shift towards WT in DAS values of pupils of isolates. The mean values of the first PC for the first generation pupils differed significantly from both ISO and WT means ($p=0.0013$ and $p<0.0001$, respectively, $n=13$ pupils). In addition, by connecting individual tutors and their pupils in this space (Fig. 2.16b), we can see that a shift toward WT occurred in every single case.

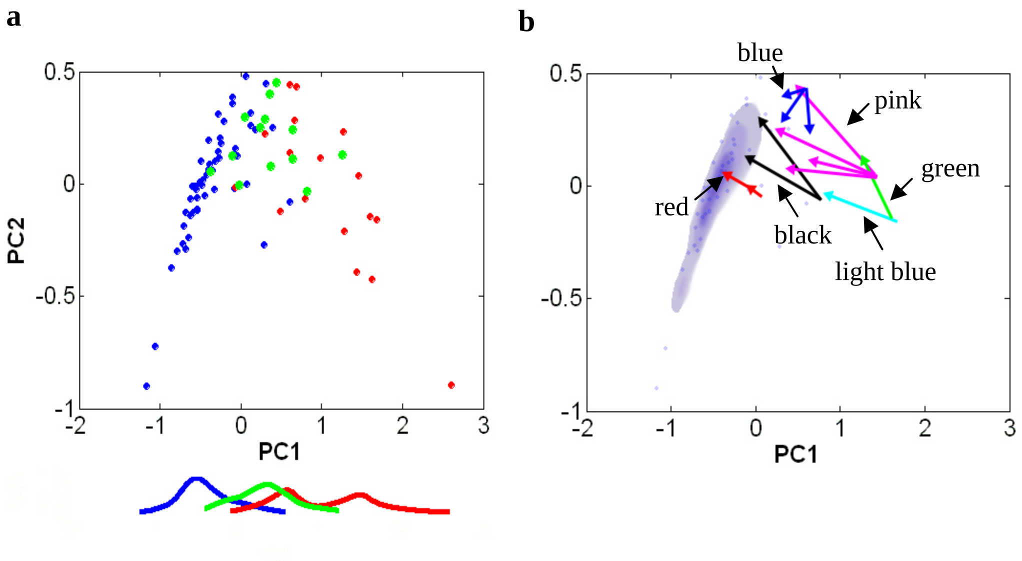


Figure 2.16 | PCA of DAS in ISO birds, WT birds and pupils of isolates. a. Each dot represents an individual ISO (red) or WT (blue) bird or pupil (green) bird. **b.** Arrows connect ISO tutors to pupils. Each arrow is a tutor/pupil pair, each color represents a different tutor.

C. Note duration ratios

Since this is the first time DAS is used as a descriptive tool, we investigated note-level changes in imitation using another feature, the distribution of note lengths. We calculated the ratio between the longest and shortest syllable within a bout. We found that this duration ratio was significantly higher in ISO tutors compared to their pupils ($p<0.01$ $n=13$, Wilcoxon sign test, Fig. 2.17).

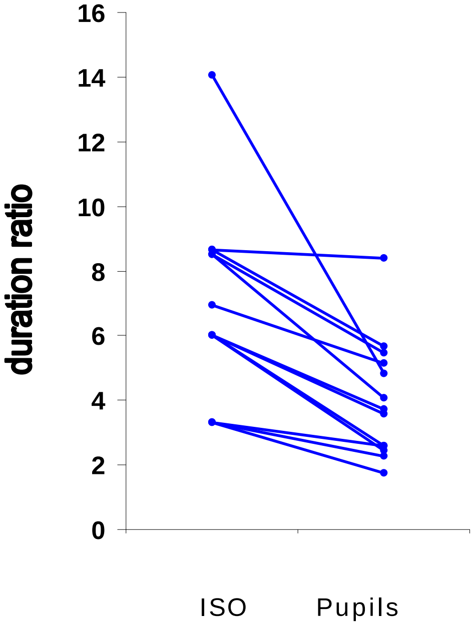


Figure 2.17 | Note duration ratios decrease in pupils of isolate tutors (n=13 ISO and pupil pairs). The left column shows the duration ratios of isolate tutors, the right column those of their pupils. In every case, the ratio decreased, showing that pupils had less variability in syllable duration.

In fact, note duration ratios of pupils of isolates fell right between those of WT and ISO birds. This is shown in Figure 2.18.

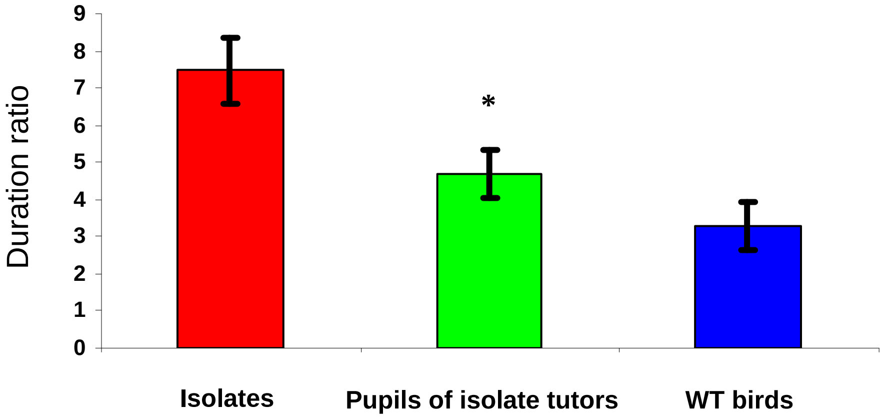


Figure 2.18 | Note duration ratios of pupils are intermediate to ISO and WT birds. Isolates (red, n=17) have the highest ratios, WT birds (blue, n=11) the lowest, and pupils of isolates are intermediate (green, n=10). Error bars indicate standard error. Asterisk above pupils indicates significant difference from ISO ($p<0.01$)

2.6 Bout-level features in pupils of ISO birds

In this section, we will examine the imitation of ISO song at the level of the bout. These include pupils' imitations of syllable abundance and song rhythm.

A. Syllable abundance

A feature that pupils consistently changed was the abundance of dominant syllables. ISO songs often contain back-to-back repetitions of the same syllable type, sometimes dominating the song in abundance (e.g., Tutor 2, Fig. 2.3). In contrast, syllable stuttering is rare in WT song. In WT songs, syllables are nearly always organized in a stable sequential order, repeating only once in a song motif. We now examine this global reorganization of syllabic structure by taking a closer look at Tutor 2 and his first pupil (Fig. 2.19). Tutor 2 sang back-to-back renditions of two syllable types (denoted as *A* and *B*). In the tutor, the abundance (relative frequency) of syllable *B* was 81%. The pupil imitated both syllables but syllable B was altered (B') and its relative frequency decreased to 19%.

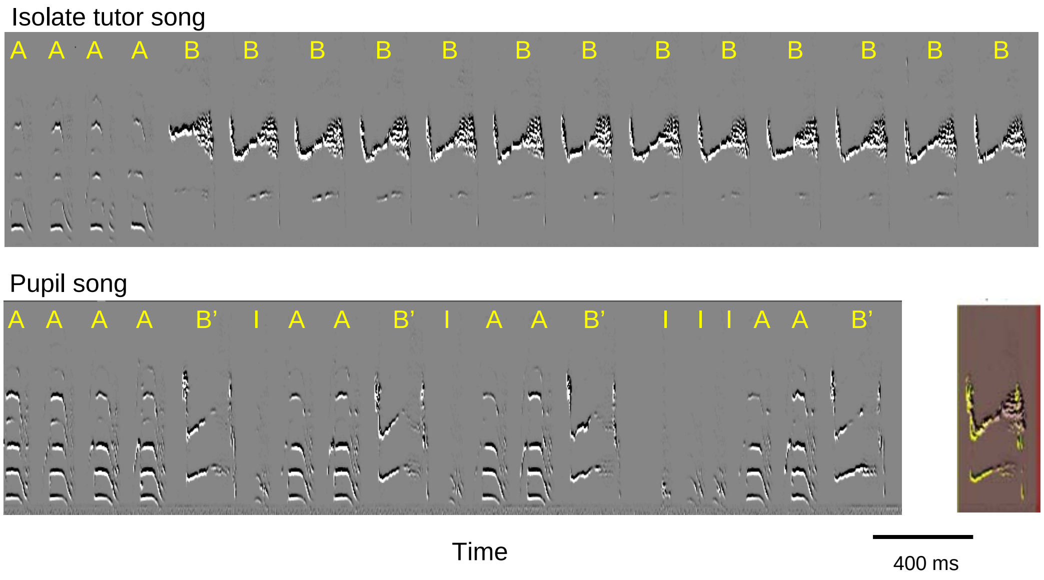


Figure 2.19 | Syntactic reorganization in pupil of isolate. Tutor 2's song consists of two syllables, both are repeated consecutively (AAAABBBBBBBBB…), but his pupil did not copy the repetitions of syllable B, rather, he constructed a motif with a sequential organization of alternating syllables (AABAAB).

Across birds we found that ISO tutors varied markedly in the relative frequency of their most abundant syllable (mean=41%, range=10-84%). These "most abundant syllables" were copied by all 13 pupils, but the relative frequencies of the same syllables in pupils' songs were significantly lower and less variable (mean=20%, range=12-30%, $p<0.01$, Wilcoxon sign test). Interestingly, when the relative frequency of the most abundant tutor syllable was 30% or lower, the relative frequency of the same syllable in the pupil song followed the tutors' values (Fig. 2.20**,** $r^2=0.77$, slope=0.85, $p=0.02$, n=6 birds). However, for syllables in the tutor song with relative frequencies higher than 30%, there was no correlation ($r^2=-0.02$, slope=0.04, NS, n=7 birds), and relative frequencies in pupil song decreased to 20-30%. Overall, the ranges of relative frequencies in pupil songs corresponded to that of WT songs, where they rarely reach 30%.

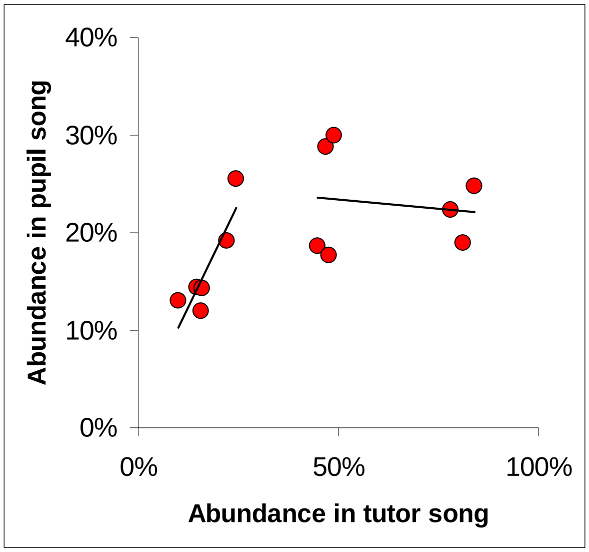


Figure 2.20 | Abundance of dominant syllable in ISO tutors' and their pupils' songs. Red dots indicate individual syllables. Syllable abundance was imitated faithfully in the normal WT range (up to 30%), but syllables dominating ISO songs too much were reduced in relative frequency to the WT level.

This shift in abundance may also indicate the formation of motifs in pupils of isolates. Zebra finch song motifs contain 3-8 syllables on average, which, when repeated only once as in WT motifs, would correspond to a 13%-33% syllable abundance for each syllable. As pupils exhibit this range, it is possible that, as opposed to their ISO tutors, they sing stable motifs.

B. Song rhythm

Some of the previously described alterations in ISO features, such as the decrease in back-to-back syllable repetitions and the shortening of syllable duration inevitably lead to changes in the global temporal structure of the song. Pupils' songs sound shorter, more structured and rhythmic. We will now attempt to quantify these rhythm changes using the technique introduced in Chapter 1.

We now turn directly to PCA, because rhythm spectra of birds singing different songs cannot be combined in histograms. In Chapter 1, we showed that ISO songs have less structure than WT songs and that when the first two principal components are plotted against each other, they occupy different regions of the distribution. We now plot the PC1 and PC2 values of pupils of ISO birds in the same graph (Fig. 2.21). As shown in the figure, most pupils (green dots) lay intermediate to the ISO (red) and WT (blue) birds. However, the separation between pupils and their ISO tutors did not reach statistical significance for rhythm.

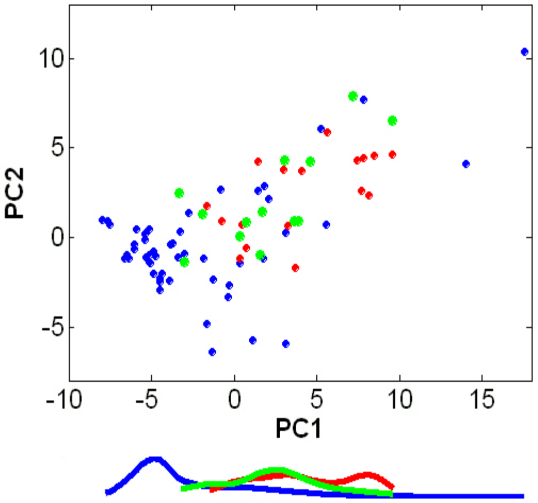


Figure 2.21 | Rhythm PCA distribution of pupils of isolates. PC1 and PC2 of rhythm of ISO (red, n=17), WT (blue, n=52) and first generation pupils of ISO birds (green, n=13). Histogram of PC1 values at the bottom. Shift towards WT features is indicated by the green dots positioned between the red and the blue dots.

Even though the effect was not significant in rhythm, it is useful to look at how the rhythm shifted between each individual tutor/pupil pair. Therefore, we present the same data with arrows drawn between tutors and pupils, as we had done for DAS and spectral features (Fig. 2.22). We can see that for some tutors (pink and black, representing Tutor 1 and Tutor 5, respectively) the rhythm shift is strong in the WT direction. Pupils of Tutor 2 and 3 (blue and red arrows, respectively) showed inconsistency in this measure, and the pupils of Tutor 4 & 6 (light blue and green arrows, respectively) shifted in the opposite direction. However, these results are very doubtful, because these last two tutors had extremely abnormal, long syllables and basically no motif at all, but their pupils sang much shorter and more stable songs (see Fig. 2.5 and 2.7). These temporal changes must have surely manifested themselves in a more structured song rhythm. In fact, to the ear, the pupils' songs sound much more normal than the tutors' songs. Consequently, the negative results are due to shortcomings and inconsistencies in our measure of song rhythm rather than a shift towards ISO rhythm.

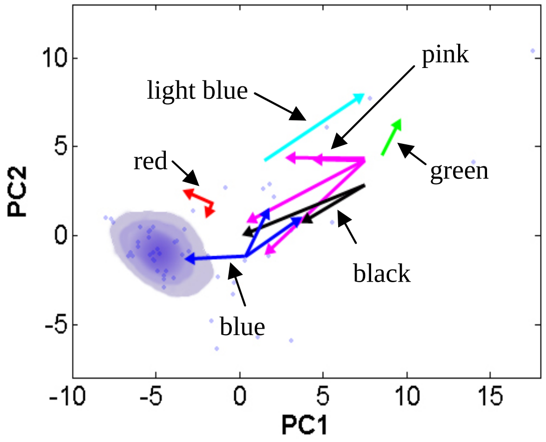


Figure 2.22 | Rhythm changes in PC1 and PC2 between ISO tutors and pupils. Arrows originate at tutors and point to pupils. Different color arrows represent different tutors. Shaded area is the density estimate of the center of WT distribution.

2.7 Conclusions

We have shown that pupils of ISO birds, while imitating their tutors, modify their songs in ways that are consistent across pupils' and tutors' songs, shifting towards WT song features. In other words, pupils learning their songs from isolates sing more normal, WT-like songs than their tutors. To statistically confirm this finding, we computed pair-wise Euclidean distances between the cumulative feature distributions of each pupil and his tutor, and between the pupil and our library of 52 WT birds. 12 of 13 pupils showed significant progression toward WT at at least one of the three timescales of song structure (FDR significance level=0.01, binomial test with n=52). Ten birds showed significance

(FDR level=0.01) at all timescales. Therefore, most birds did shift towards WT features regardless of what song feature is considered.

The implications of this shift in pupils of isolates are great, because it provides evidence to the existence of innate biases that shape song imitation and may form the basis of evolutionary changes, such as dialect formation and preservation. Perhaps most interestingly, these biases surfaced immediately upon providing a tutoring environment but not without it. Isolates do not show these biases when they develop their own songs, but their pupils, even though they are confined in some sense by the abnormal tutor song, are able to employ them to shape and modify their imitations of the tutor songs.

We find it remarkable that the syllables invented by the isolates, who still use auditory feedback to develop these syllables (Price 1979), are not subject to the biases that we see in their pupils. It appears as if the isolate bird has “no problem with” hearing himself singing a bizarre scratchy syllable, that pupils of isolates show no biases against imitating bizarre ISO syllables, and yet when the pupil copies these syllables, he shifts their features toward WT. We therefore interpreted the shift toward WT songs observed in the pupils as an outcome of *imitation biases*. Namely, biases that are associated with the imitation of songs – not with production per se, and not during the selection of song model (which would have resulted in selective imitation of some syllables over others).

Imitation biases are likely innate. We will refer to these biases as innate hereafter, although we cannot exclude the possibility that some involvement of early social experience with the mother affected the development of such biases.

It was surprising that we saw such a strong shift towards WT features in pupils of isolates. There are instances of rapid dialect changes in the wild, but they usually involve the addition of new song elements that become “fashionable” in a population. Such processes have been described in Puget Sound white-crowned sparrows (Chilton & Lein 1996) and humpback whales (Noad et al. 2000). What is very interesting in our results is that pupils did not add or delete sounds much, but rather, manipulated the existing cultural variant and fit it to an “innate idea of what it should be like” – which is the WT structure.

We will next explore what happens to these songs in a recursive tutoring paradigm, where the pupils of ISO tutors become tutors themselves and so on. We will track the evolution of songs over a few generations of song learners.

Chapter 3. Evolution of ISO song over multiple generations

Background & Rationale

In the previous chapter we showed that the imitation of ISO songs by young zebra finches is more similar to WT songs than the song of the ISO tutor. We observed a shift towards WT song features at three timescales of song structure: in spectral features (frequency modulation, amplitude modulation and goodness of pitch), note-level features (such as note & syllable length and duration of acoustic state), and bout-level features (syllable abundance and song rhythm). We saw that rather than preferentially imitating WT-like syllables, pupils changed the ISO-like syllables to be more WT-like, and we call this effect imitation biases to indicate that these biases must occur during the process of imitation (as opposed to production biases, or selective imitation). The biases that drove the changes were quite similar across tutors and across pupils.

The next question that arises is what happens to the song beyond the first generation. Perhaps, the first generation birds have revealed the full extent of innate biases, so the ISO song has been changed as much as it can be towards WT. In this case, the changes would have been a reaction to an overwhelmingly abnormal song stimulus, but pupils of first generation birds would imitate faithfully. However, we saw that the songs of the first generation birds are still as different from WT songs as from ISO songs, so it is possible that the progression towards WT-like features will continue if we let later generations of pupils learn the songs. In the present chapter, we will follow the subsequent evolution/development of the songs as they are passed down over generations of song learners and try to answer the question whether the progression towards WT-like features continues over multiple learning generations. As noted in the Introduction, cumulative cultural evolution is thought to occur in birds, but there is no direct empirical evidence. If it is true that the progression from ISO to WT takes multiple generations, then, by definition, song culture is cumulative – namely, the WT singing culture we observe cannot be achieved in a single generation, but it can be achieved by recursive learning.

Methods

Upon reaching adulthood and song crystallization, we used 4 of our pupils from the ISO tutor/pupil experiment to train young males individually. When adult, these first generation pupils become the tutors of the next generation pupils, and so on, recursively (Fig. 3.1). As before, pupils were assigned randomly to their tutors (minimizing genetic relatedness). The isolate tutor and his pupil were kept together for 90 days in a sound isolated chamber. Subsequently, we removed the tutor and recorded the pupil's song and compared it to that of his tutor. Following this, we placed a 30-day old juvenile male in the sound chamber housing the now mature pupil who became song tutors for the next generation pupils.

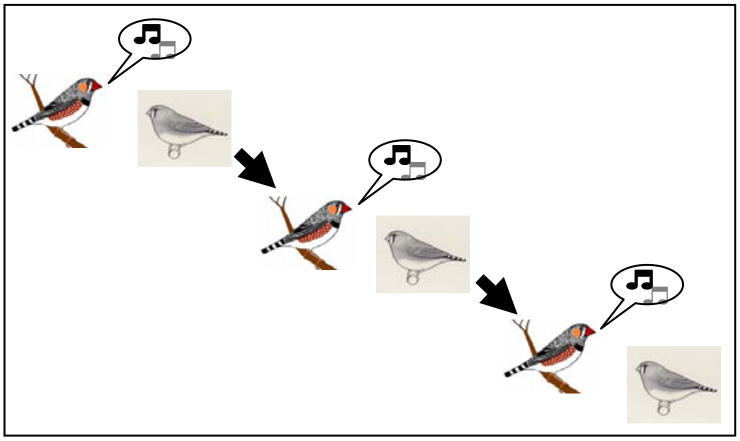


Figure 3.1 | Schematic diagram of multigenerational tutoring paradigm. The pupil (grey birds), upon reaching adulthood, becomes the tutor (colorful birds) of another juvenile. The tutoring continues recursively over several learning generations.

Table 3.1 presents the tutor and pupil identities and generation numbers for each bird in the multi-generation experiment. Note that the lineages and generations indicate recursive learning generations and not genetic relations.

	Tutoring Lineage 1	Tutoring Lineage 2	Tutoring Lineage 3	Tutoring Lineage 4
ISO Tutor	19 (Tutor 1)	1211 (Tutor 2)	1238 (Tutor 3)	1249 (Tutor 5)
1st generation pupil	1248	1402	1342	1439
2nd generation pupil	1326	1514	1571	1558
3rd generation pupil	1374	1606		
4th generation pupil	1535			
5th generation pupil	1621			

Table 3.1 | Tutoring lineages in multigenerational tutoring experiment.

Results

3.1 Visual assessment of imitation across tutoring generations

As before, we begin by examining the outcome of multi-generational training by inspecting the sonograms and comparing tutor songs to all of their successive pupils' songs.

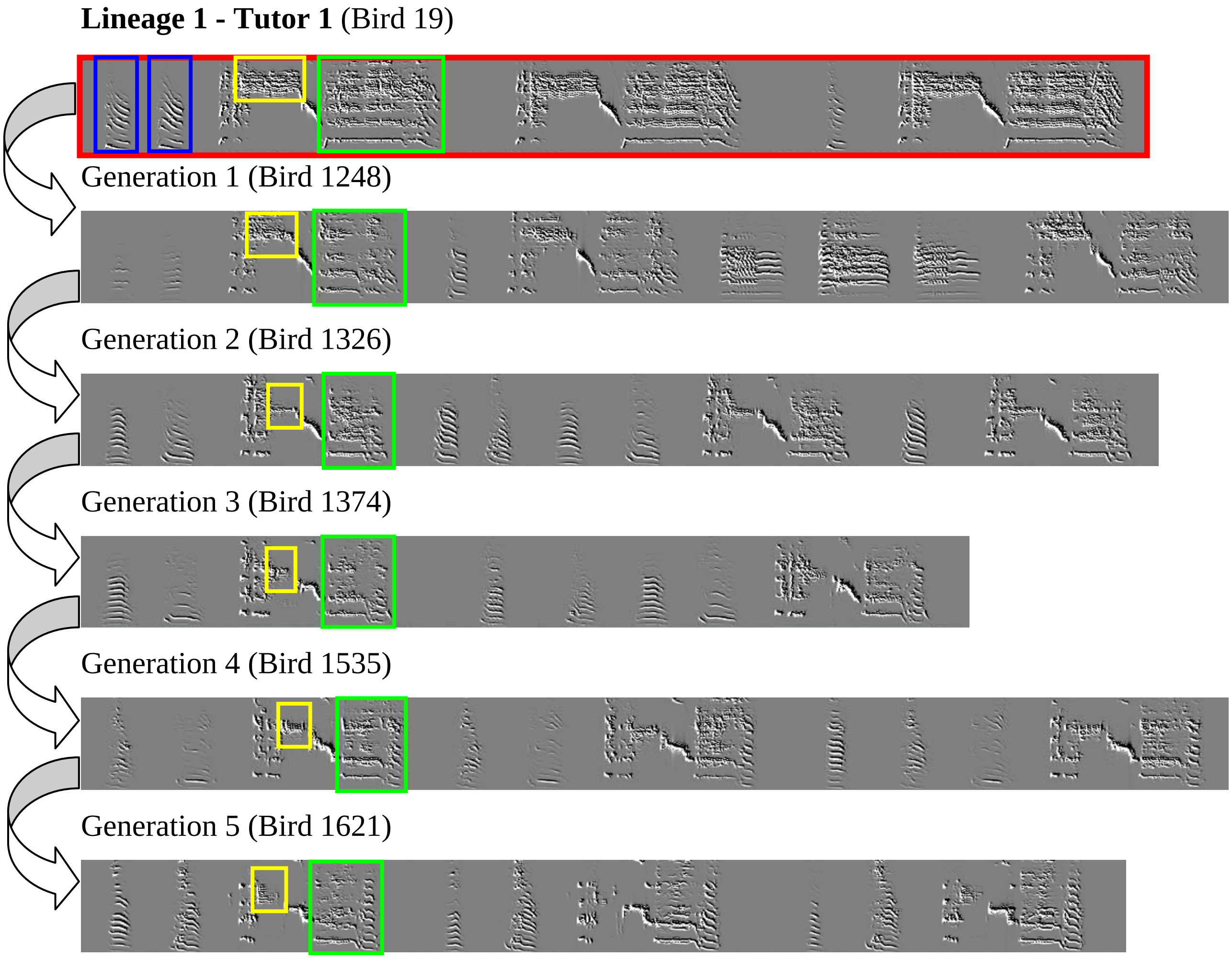


Figure 3.2 | Tutoring lineage of ISO Tutor 19. Notes that were altered are marked with different colored rectangles (blue, yellow and green in order in ISO song).

Following the song of Lineage 1 (Fig. 3.2) over a few generations reveals that much of the structure of the original tutor song was preserved by the pupils. Interestingly, the direction of vocal changes that the first generation pupil made persisted over generations and therefore, changes accumulated in the succeeding generations. For example, the copies of the long and monotonic ISO note (green rectangle) kept decreasing in duration in later generation pupils and its spectral structure was modified and differentiated gradually, so that by the fifth generation pupil, the note became shorter, but also differentiated into two distinct notes, one broadband and the second with pure harmonic structure. Similarly, the yellow note became much reduced in duration, bandwidth and in amplitude over generations of learners, and by the fifth generation, there was only a trace of the original syllable.

There were bout-level changes that were made to this song, as well. The bout got stretched out by the introduction of other, short (introductory-like) syllables between the renditions of the long syllable. The second generation pupil sang some medium duration modulated call-like syllables between the motifs. These syllables were not improvised, as they can be found in the isolate tutor’s repertoire, although at much lower frequencies. His pupil sang them in nearly every bout. These syllables didn’t become prevalent in the later generations pupils’ songs.

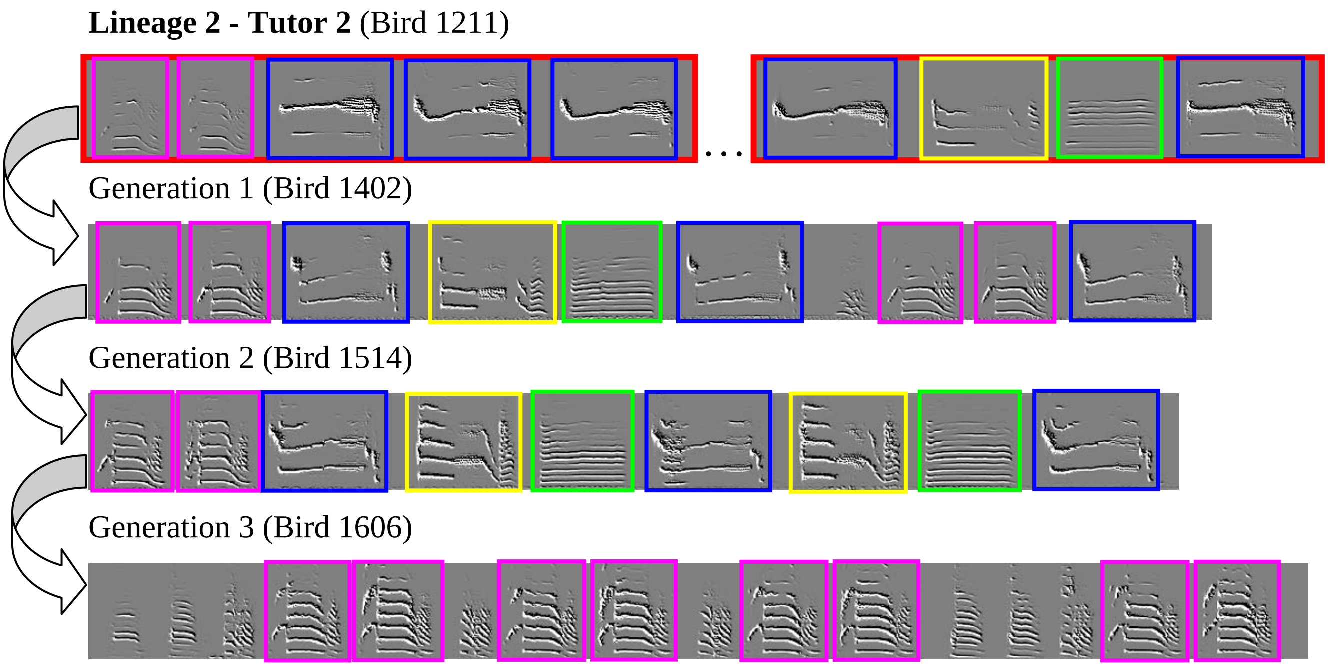


Figure 3.3 | Tutoring lineage of ISO Tutor 1211. Important syllables are marked in different colored rectangles (pink, blue, yellow and green in order).

In Lineage 2 (Fig. 3.3), the isolate tutor sang a very long, high-pitched syllable as the dominant element in his motif (blue rectangle). This song was very unusual both because of the spectral features of the dominant syllable and because it was repeated consecutively (15 times on average) by the ISO tutor. Two other syllable types of similar duration (in yellow and green), sung only once, were nested among these repetitions. Interestingly, the first generation pupil did not imitate the repetitions and constructed his motif out of the three long syllable types, singing them serially once and ending with the first one. Another syntactical reorganization happened in the second generation, when the song went from one rendition introduced by two renditions of the pink syllable, ppABCA, to ppABCABCA. This motif repetition and bout lengthening is reminiscent of the syntax changes that take place during development in an individual bird. The generation 3 pupil completely omitted the long syllables and sang a simple song whose spectral features were WT-like. This was one of the very few instances of selective imitation, where the pupil only imitated the more WT-like syllables and did not imitate the abnormal syllables at all.

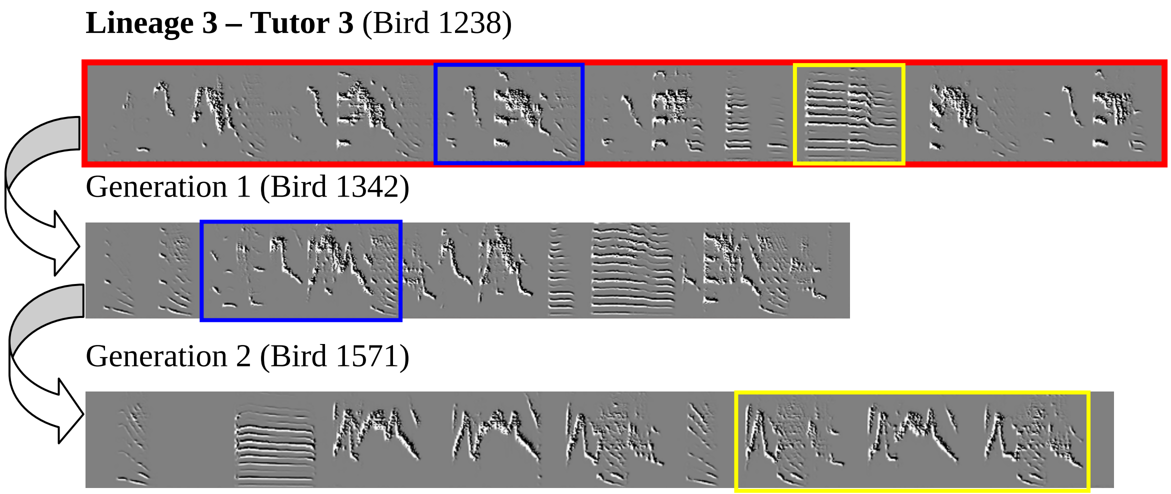


Figure 3.4 | Tutoring lineage of ISO Tutor 1238. Important syllables are marked in different colored rectangles (blue and yellow).

Tutor 3 (Fig. 3.4) sang a complex syllable (blue rectangle) that, based on its acoustic features and fast transitions, could be classified as a WT song. However, the syllable repetition is highly unusual and the stuttered harmonics (longest in yellow) that follow these repetitions are also abnormal. The repetition rate decreases in the song of the first generation pupil and the blue syllable becomes even more complex and differentiated so that it is not repeated exactly the same way each time. The second generation pupil progresses with the differentiation and breaks up the long syllable into two independent types. We still see some syllable repetition, but most of the time alternating syllables are sung in serial order (yellow).

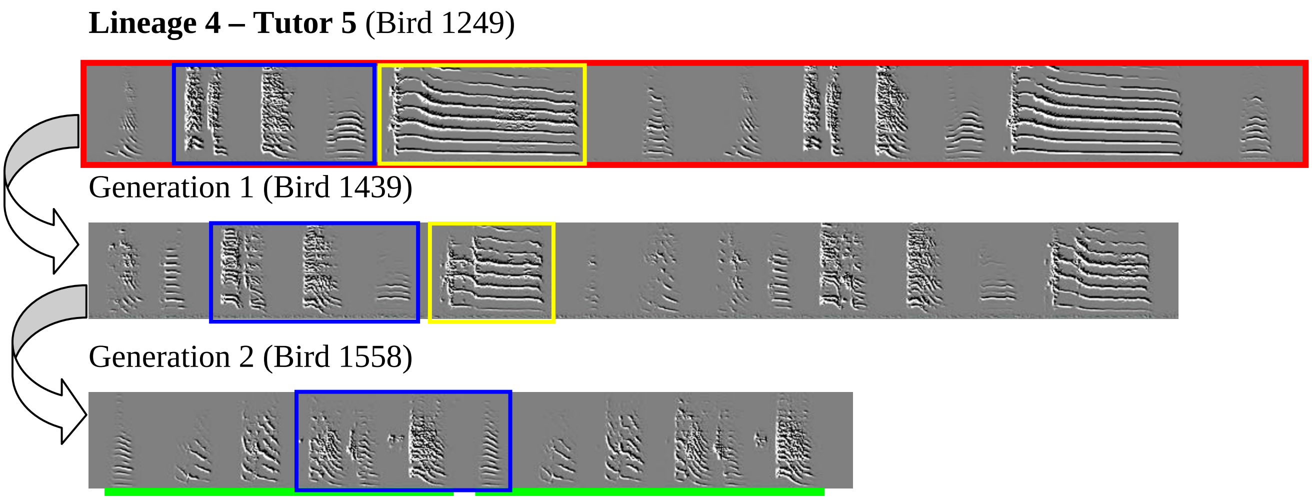


Figure 3.5 | Tutoring lineage of ISO Tutor 1249. Important syllables (yellow) or groups of syllables (blue) are marked in different colored rectangles.

Tutor 4 (Fig. 3.5) sang a simple song with some short, introductory-like notes (blue rectangle) followed by a short and a very long harmonic syllable (yellow rectangle). The first generation pupil imitated everything with high accuracy except he shortened the harmonic, but the second generation pupil omitted the harmonic and changed the structure of some of the short notes (first and last note in blue rectangle). In addition, he introduced a short, high-pitched note in between the short syllables. Due to the shortening of silence intervals, the song motif in the 2nd generation pupil's song (underlined in green) is short and stable like WT motifs.

Summary of subjective inspection

Overall, the biases that surfaced in the first generation pupils were continued by successive pupils. Long ISO syllables were shortened further, spectral complexity and stability increased further and differentiation into notes continued over multiple learning generations.

Next, we present a numerical summary of the imitation accuracy and syllable identities in tutoring lineages.

3.2 Imitation statistics

Tutor	**19**	**1248**	**1326**	**1374**	**1535**
Pupil	**1248**	**1326**	**1374**	**1535**	**1621**
Tutor: complex	1	1	1	1	1
Pupil: complex	1	1	1	1	1
% copied	100	100	100	100	100
% invented	0	0	0	0	0
Accuracy	81	86	93	81	73
Tutor: call-syll	0	1	1	0	1
Pupil: call-syll	1	1	0	1	1
Tutor: rare syll	2	0	0	0	1
#calls copied	3	1	0	0	1
#calls invented	0	0	0	1	0

Tutor	**1211**	**1402**	**1514**	**1238**	**1342**	**249**	**1439**
Pupil	**1402**	**1514**	**1606**	**1342**	**1571**	**1439**	**1558**
Tutor: complex	3	3	3	4	3	2	2
Pupil: complex	3	3	2	3	2	2	2
% copied	100	100	25	75	66	100	100
% invented	0	0	0	0	0	0	0
Accuracy	88	88	92	81	66	90	84
Tutor: call-syll	1	1	2	3	3	3	3
Pupil: call-syll	2	2	1	3	2	3	1
Tutor: rare syll	2	0	0	0	0	0	0
#calls copied	2	2	1	0	2	3	1
#calls invented	0	0	0	0	0	0	0

Table 3.2 | Imitation of isolate tutors over multiple generations of pupils. Columns show the individual statistics for every tutor and his pupil.

Table 3.2 shows that imitation level was generally high, with fairly high accuracy. Innovation was low and complex syllables were imitated in almost all cases. More birds invented call-like syllables than complex syllables. Sometimes the accuracy decreased again in later generations, for example in Tutor 19’s lineage, because at first the birds made changes in duration but when that stabilized, spectral modifications were made. This implies that the changes at different timescales are not necessarily occurring simultaneously.

3.3 Progression toward WT continues in multiple learning generations

We tested for additional multigenerational progression towards WT song features using the first Principal Component, as in the previous chapter. Mean PC1 values were compared across groups for spectral features, duration of acoustic state and rhythm. Our groups are now ISO birds, 1st generation pupils, later generation pupils and WT birds.

We first tested if later generation birds progressed toward WT features beyond the level achieved in the first generation pupils. Direct comparisons across first and later generation pupils reached significance only for DAS ($p=0.02$), but multigenerational comparisons suggest further progression towards WT for all song traits.

For spectral frame features, we found that the PC1 of spectral features changes monotonically towards WT over generations. Its mean values for ISO, first generation, later generations, and WT songs were 1.3, 0.3, 0.03, -0.4 respectively. First PC values for later generation songs were significantly different from ISO song ($p<0.005$, t-test, $n=8$ for later generations) but not from WT songs ($p=0.17$). In other words, later generation birds were already within the WT cluster, whereas first generation birds were still significantly different from WT. A close inspection of the result suggests that the lack of significance in direct comparison is because some of the first generation pupils reached WT levels in spectral features, namely, there is a saturation effect.

For DAS, direct comparison reached significance, and indirect assessment gave similar results to that of spectral frame features, namely, the PC1 values also decreased monotonically with generations: 1.1, 0.3, 0.02, -0.3. However, higher generation songs were significantly different ($p<0.01$) from both WT and ISO, indicating that WT approximation was not complete even in the later generations.

For rhythm, results were not significant even in the first generation, and yet, indirect assessment suggest further progression over generation. PC1 values also decreased monotonically with generations: 4.1, 2.2, 1.4, -2, and differences from WT and ISO were marginally significant ($p=0.02$, 0.056 respectively). Statistical tests are further described in Appendix III.

After showing that progression towards WT song features over multiple generations continued in later generation pupils, we attempt to judge qualitatively if similar biases to those we found in first generation pupils can also be seen in the later generation pupils. Again, we investigate these changes at three different timescales of song structure: spectral features, duration of acoustic state and song rhythm.

3.4 Multigenerational transitions in spectral frame features

As noted earlier, three spectral features showed significant difference in distribution across WT and ISO: AM, FM and goodness of pitch. Figure 3.6 shows the probability distributions of these features for all of our first and multigenerational pupils. Red curves represent ISO birds ($n=17$), blue curves represent WT birds ($n=52$) and green curves pupils ($n_{Gen1}=13$, $n_{Gen2}=4$, $n_{Gen3}=2$, $n_{Gen4}=1$, $n_{Gen5}=1$). The top panels include WT birds as well for reference. The other panels only include the ISO Tutors (red, $n=6$) and the respective pupils. These plots demonstrate that in all three spectral features there is a shift towards WT distributions in first generation birds and that this shift is maintained or continued in later generation pupils. For example, for AM (Fig. 3.6a), we see in the top panel that WT birds (blue curves) have a lower peak than ISO birds (red curves). Green curves have a lower peak. Then in later generations this peak remains consistently lower than the tutors’ peak, and up until the 5th generation, there is even a slight decrease in the height of the peak. There is a small backwards tendency in AM and FM in the 5th generation, but this should not be taken too seriously as the sample size in the 5th generation is 1. In FM (Fig. 3.6b), similarly to AM, there is a downwards shift in very low values in pupils towards the WT distribution. We mentioned that ISO birds have lower FM typically, so this shift indicated that pupils’ songs do not contain so many low FM sounds. In goodness of pitch (Fig. 3.6c), it seems that most or all of the effect takes place in the first generation, but the curves of later generation birds (green curves) tend to be lower and to the left of ISO tutors (red curves).

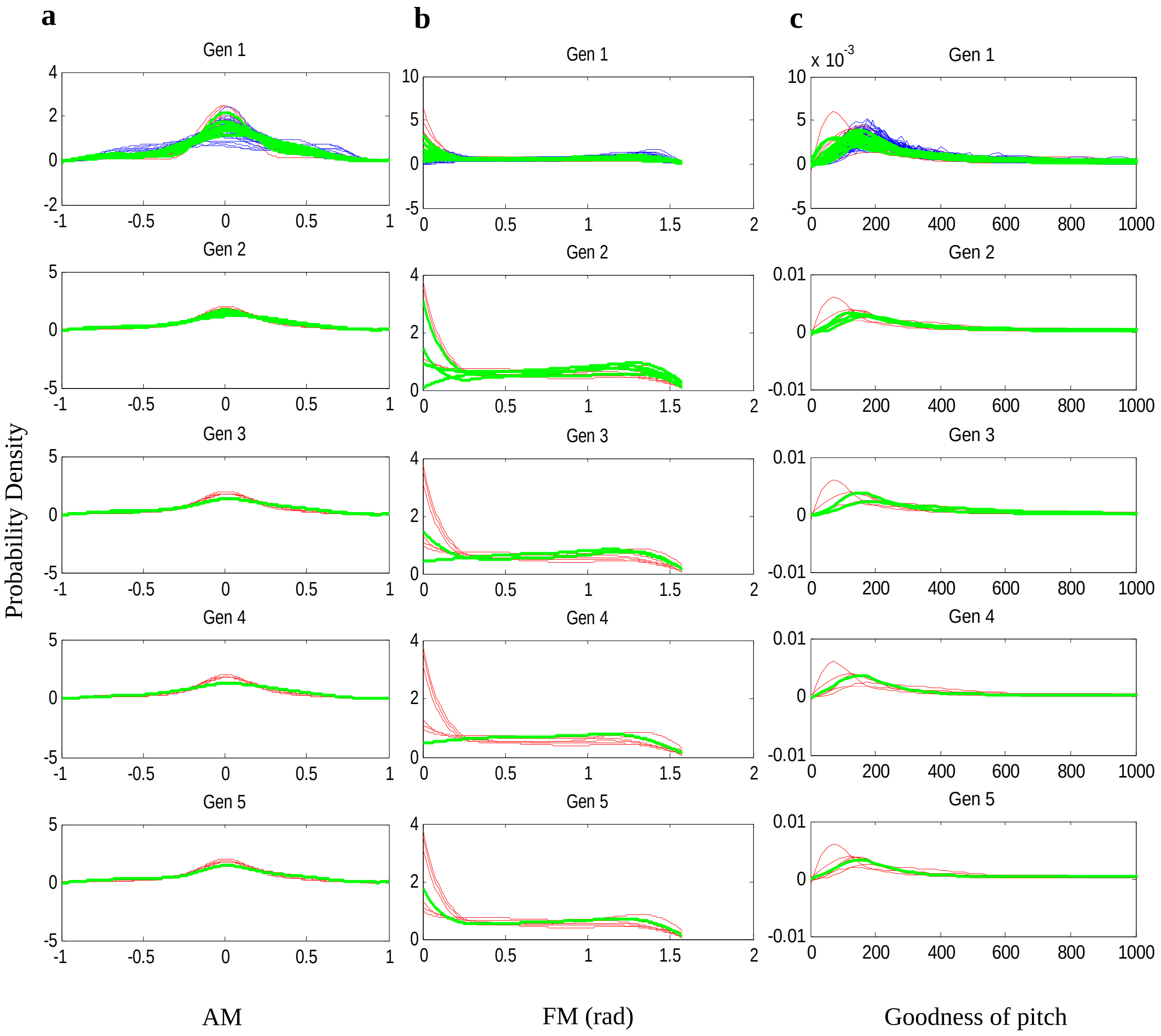


Figure 3.6 | Probability distributions AM (a), FM (b) and goodness of pitch (c) in songs of multiple learning generations. Top panels contain all ISO (red, n=17), WT (blue, n=52) and first generation birds (green, n=13). Lower panels contain ISO Tutors (red, n=6) and later generation birds (green, n_{Gen1}=13, n_{Gen2}=4, n_{Gen3}=2, n_{Gen4}=1, n_{Gen5}=1).

Considering all the spectral features combined, we get a more complete picture of the multigenerational progression towards WT song. From the probability distributions, we calculated cumulative distributions and from those, we performed PCA. PCA plots provide a good visual tool to investigate multigenerational progression. Figure 3.7 shows the first two principal components (PC1 and PC2) with all of our ISO, WT and tutored birds, and the arrows mark the first two generation pupils. As before, arrows point from the tutor to the pupil. The numbers near the origin of the first arrow is the ISO Tutor's name. The second generation pupils in all tutoring lineages continued to shift away form the original ISO tutors in spectral features.

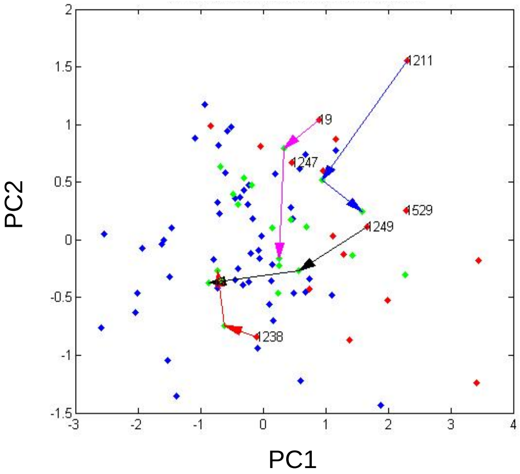


Figure 3.7 | PCA distribution of combined spectral features of first and second generation pupils of isolates. Arrows show that second generation pupil continue the progress away from the ISO tutors.

The representation above is somewhat crowded, which makes it difficult to visualize progression over several generations. To simplify the image and enhance visualization, in addition to individual WT birds, we present the WT density distribution as a cloud. We can now see clearly how the WT distribution is shaped and where it is centered without showing individual birds (Fig 3.8). We can now look at all the multigenerational pupils with reference to WT distribution to see if the continued shift away from ISO tutors also represents a progression towards WT. Fig. 3.8 shows that the progression towards WT indeed continues and does not stop until the pupils' are well within the WT distribution. In the case of our first tutoring lineage (ISO Tutor 19, pink arrows), the progression stops after the second generation, but by this time, the song is nearly in the middle of the WT cloud. Although the arrows still show movement, it seems to hover randomly around the WT distribution center. For the other songs, progression continues for as long as we have data. Based on PC1 mean values, first generation birds are significantly different from both ISO and WT (p=0.0126 and 0.018 respectively), but later generation birds only differ significantly from ISO (p=0.0045) but not from WT.

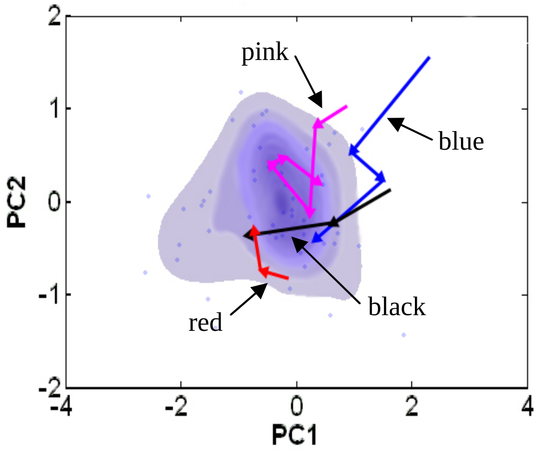


Figure 3.8 | PCA distribution of combined spectral features of multigenerational tutoring lineages. Arrows penetrate deeply in the WT distribution cloud (in grey). Lineage 1: pink arrows, Lineage 2: blue arrows, Lineage 3: red arrows, Lineage 4: black arrows.

3.5 Multigenerational transition towards WT in note-level features

We next investigate alterations in syllable durations over multiple generations of pupils. We will use the durations of acoustic state and note duration ratios to assess multigenerational tendencies.

A. Duration of acoustic state

The sonograms presented in section 3.1 suggested to us that the copies of long ISO syllables were shorter in first generation pupils, and that this trend continued over multiple generation pupils. This suggests that either the pupils compressed (time-warped) stationary syllables only, or shifted acoustic states more rapidly regardless of the copied syllable being stationary or not (namely it would apply to “complex” syllables, too). We now look at the progression of the duration of acoustic states to test two hypotheses:

1. Does DAS decrease over generations toward WT?
2. Is the change in DAS due to selective shortening of longer states?

We start by plotting cumulative histograms averaged across birds in each learning generation (Fig. 3.9b). As a reference, cumulative histograms for all ISO (red curves) and WT (blue curves) are shown in Fig. 3.9a. The gap between the ISO Tutors and the pupils increases in every generation, which indicates that histograms shift in the WT direction and that duration of acoustic state values are lower in later generation pupils.

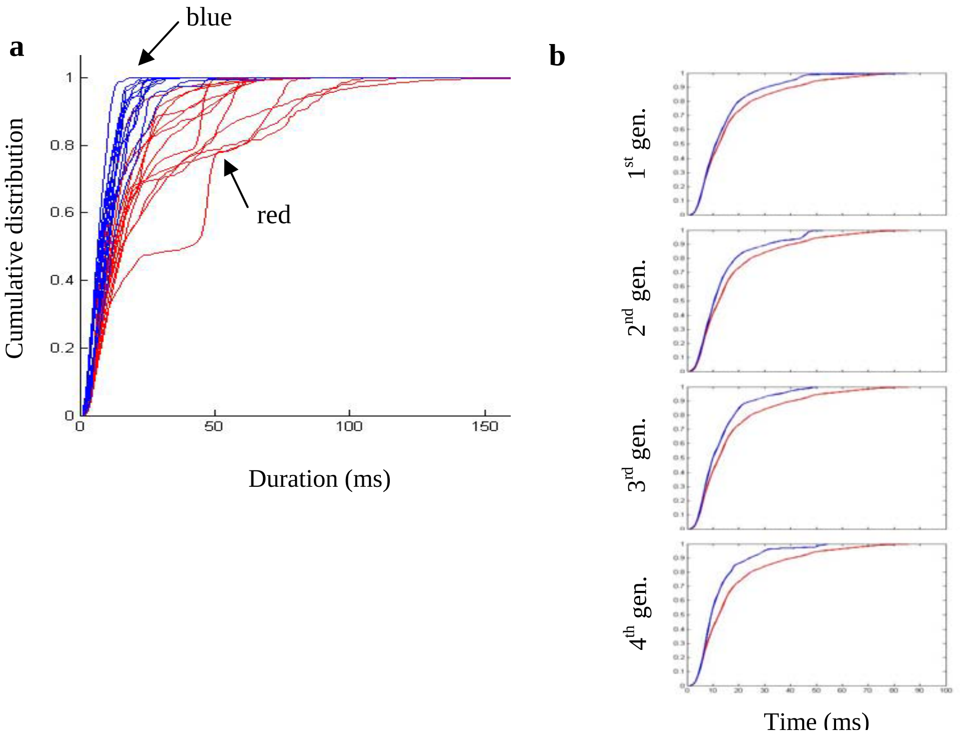


Figure 3.9 | Cumulative histogram of duration of acoustic state in multiple generation birds. a. Cumulative histogram of all ISO (red, n=17) and WT (blue, n=52) birds. **b.** ISO tutors (n=6) are shown in red in every panel (always on bottom), pupils in blue (always on top); panels from top to bottom: first generation pupils (n=6), second generation pupils (n=4), third generation pupils (n=2), fourth generation pupil (n=1).

We see the upward shift in Fig. 3.9, but the magnitude of this event is difficult to judge visually. The Kolmogorov-Smirnov (KS) statistic is an estimate of the difference between any two distributions, and it is sensitive to differences in means, in variances, and in slopes. We calculated the KS-statistic for the cumulative histogram of every bird to estimate their distance from the WT histogram. Figure 3.10 shows these values for every bird in every tutoring lineage (blue lines) and for the mean across lineages (black line). As shown, with the exception of one lineage, where there is further decrease in the KS distance from WT in generation 2, and the distance decreases even more in later generations, where it seems to asymptote. The exception is bird 1558 in Lineage 4. The original ISO Tutor was bird 1249 (Fig. 3.5), who sang very short, fairly simple notes followed by a very long call. Bird 1558 did not imitate the call-like syllable and as a result, his song only contains very short notes. Because WT zebra finch songs tend to contain medium-length harmonic notes, it is likely that this bird falls farther from the WT distribution due to its notes being shorter than the typical WT range.

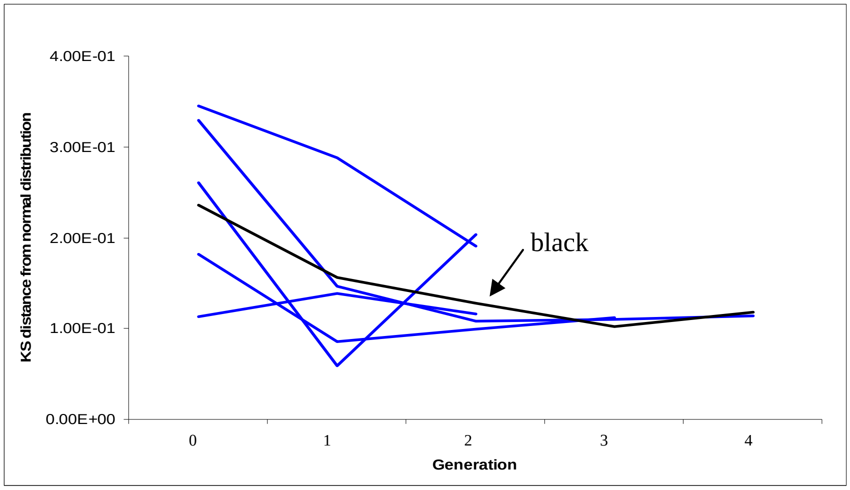


Figure 3.10 | KS statistic for WT and multigenerational pupil cumulative distributions of DAS. Blue lines (pupils in individual tutoring lineages) indicate a decreasing distance between the pupils and WT birds as generation numbers increase. Black line represents mean across lineages.

Finally, we examine the PCA for the duration of acoustic state. Figure 3.11 shows the PC1 and PC2 values for all the WT (blue, n=52) and ISO (red, n=17) birds and pupils (green, n=19) for the 4 multigenerational tutoring lineages. The first two generation pupils are connected by arrows to their tutors. The original ISO Tutor's name is written next to its data point. We can see that for Tutor 19, 1211 and 1249, the progression toward WT is fairly strong as the pupils shift away from the tutors in a general direction of the blue dots. The pupils of 1238 do not seem to be moving in the direction of the WT distribution.

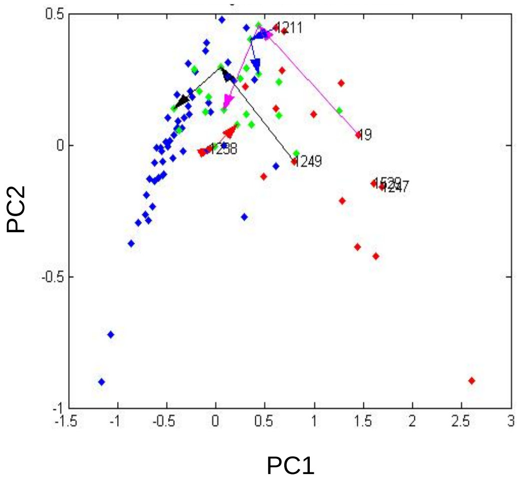


Figure 3.11 | PCA distribution of DAS of first and second generation pupils of isolates. Arrows show that second generation pupil continue the progress away from the ISO tutors.

Figure 3.12 shows all the pupils of all four tutoring lineages using the blurring visualization method discussed earlier. Over multiple generations, we see clear progression towards the center of the WT distribution. In Lineage 1 (ISO Tutor: 19, pink arrows), after generation 2 the trajectory hovers randomly within the WT area, similar to how it happened in the spectral features PCA. Overall, we see that arrows are shorter as they approach the WT center. The only lineage that didn’t approximate WT distribution is Lineage 3 (red) where ISO Tutor 1238 sang a complex song with lots of fast transitions, so that the starting point was already close to the WT center. Although the second generation pupil sang more stable syllables, there seems to be less of a variety of acoustic state durations, such as short notes and medium-length notes. We calculated the mean values for PC1 in ISO, first generation, later generation and WT birds. Both first and later generation birds differed significantly from ISO ($p=0.0013$ and $p<0.0001$, respectively) and WT ($p<0.0001$ and $p=0.0089$, for first and later generations birds, respectively), but later generations birds were also significantly different from first generation birds ($p=0.0467$).

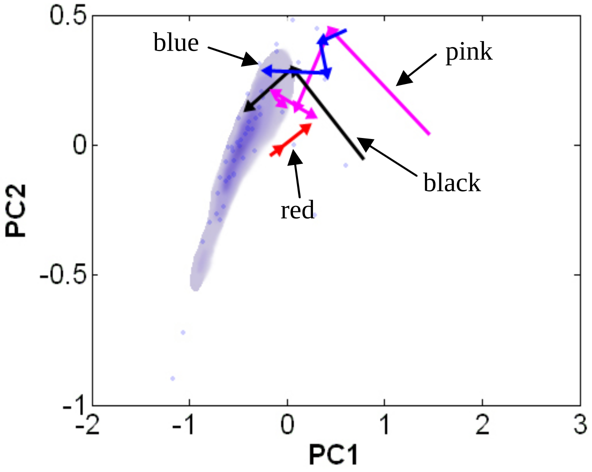


Figure 3.12 | PCA distribution of DAS of multigenerational tutoring lineages. Arrows penetrate in the WT distribution cloud (in grey) in 3 out of 4 cases. Lineage 1: pink arrows, Lineage 2: blue arrows, Lineage 3: red arrows, Lineage 4: black arrows.

We have shown that DAS values do indeed decrease over multiple generations. We now turn to measures of note duration to test our second hypothesis, that is, whether the decreases in DAS were due to the shortening of long notes.

B. Note duration ratios

Our other note-level measure to investigate the multigenerational progression of songs is note duration ratios. We calculated them for each bird and plotted the mean ratios in Figure 3.13. The multigenerational pupils (green) are shown between the ISO (red, n=17) and the WT (blue, n=11) values. In increasing generations, the note duration ratios decrease and become more similar to WT values. This suggests that DAS decreases were due to the shortening of long notes.

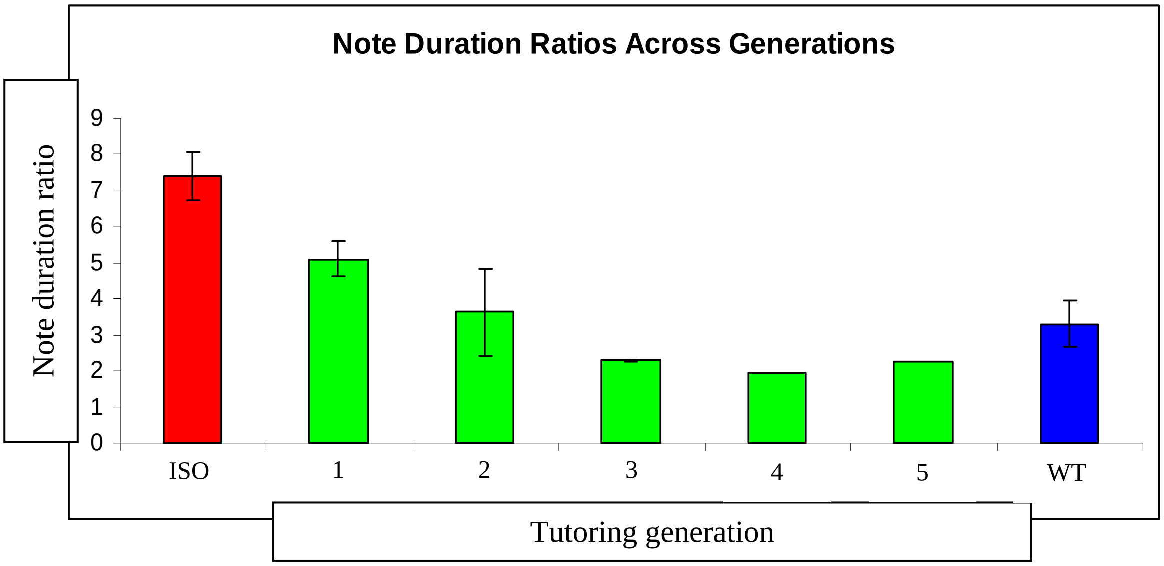


Figure 3.13 | Note duration ratios in multiple generations. ISO birds (red, n=17) and WT birds (blue, n=11) flank the pupils (green bars, n_{Gen1}=13, n_{Gen2}=4, n_{Gen3}=2, n_{Gen4}=1, n_{Gen5}=1). Error bars indicate SEM. In the 4th and 5th generation no error bars are shown, because the sample size is 1.

3.6 Transitions in song rhythm over multiple generations

In Chapter 2, we saw that the song rhythm of pupils of isolates showed a tendency to be more WT-like, but there were some exceptions. Some of these exceptions were likely due to a shortcoming in the method rather than the pupil rhythms becoming less structured. Looking at the multigenerational trajectories (Fig. 3.14), we see similar inconsistencies. The rhythms of multigenerational pupils in three out of the four lineages (Lineage 1, pink; Lineage 2, blue; Lineage 3, red) shift back towards WT, and only one lineage (Lineage 4, black) progresses robustly towards the WT distribution, although it doesn't penetrate by the second generation. PC1 means did not differ from ISO for either first or later generation birds, but both were significantly different from WT (p=0.0033 and 0.017, respectively).

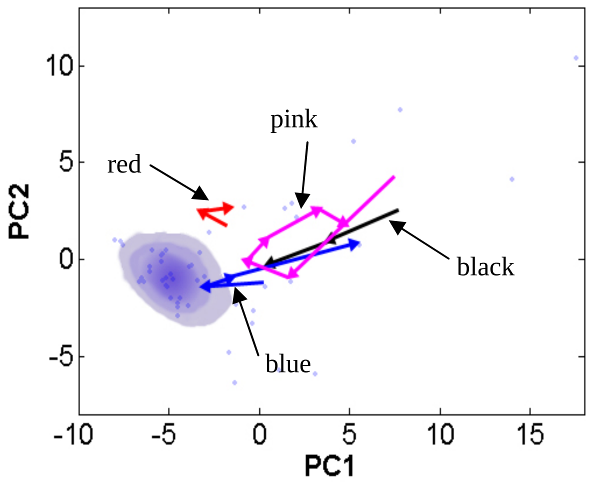


Figure 3.14 | PCA distribution of rhythm spectra in multigenerational tutoring lineages. Arrows penetrate in the WT distribution cloud (in grey) in 3 out of 4 cases. Lineage 1: pink arrows, Lineage 2: blue arrows, Lineage 3: red arrows, Lineage 4: black arrows.

3.7 Conclusions

In this chapter we showed that the progression towards WT features continues over multiple tutoring generations in a recursive learning paradigm. The strategies that the first generation birds employed, namely, imitating WT-like and ISO-like syllables both, but modifying the latter during imitation, were also used by multigenerational birds. The changes were to a lesser degree, but this was due to the fact that first generation pupils sometimes reached WT range feature distributions. In cases where this did not happen, pupils of later learning generations modified more. Overall, within 3-4 generations songs became WT-like at three timescales of song structure: spectral features, acoustic state duration and note length, and song rhythm. We found the weakest effect in song rhythm, and the greatest effect in spectral features.

Chapter 4. Cumulative cultural evolution in an isolated colony

Background & Rationale

The first time the question of multigenerational progression towards WT surfaced, we set out to answer it by recreating the “island effect” in a semi-natural condition, by starting a colony with an ISO male and observing how his offspring imitate his song. This experiment took lots of effort and time, and while in the initial stages of gathering data, we started the one-to-one tutoring experiments described in Chapter 2 and 3. We left the presentation of our colony experiment last, because this experiment could only be done once, and therefore, the outcome cannot be judged statistically. On the other hand, the semi natural colony experiment allows us to judge if, and to what extent, the phenomenon we observed in an impoverished environment, might be similar to what may occur in a more natural social environment.

In a colony, several additional social factors can affect the outcome of song imitation. Female zebra finches, when raised in the presence of adult males, show a preference for WT song over ISO song (Searcy et al. 1985). Therefore, the females might mate selectively with males who sing certain songs, and hence bias the development of song culture in a direction different than the one we see in the one-to-one training. Second, the females might potentially train the males to sing in a certain manner (action-based learning, West & King 2000, Nelson & Marler 1994). Another factor that might change the dynamics in a colony setting is social inhibition of song imitation by siblings (Tchernichovski 1998). For example, in a clutch with multiple male offspring, imitations diverge and typically only 1 or 2 brothers produce a good imitation. Of course, there are many other social forces in a colony that could potentially alter song learning and song culture evolution. Looking at the development of song culture in a semi-natural setting cannot tell us which one of these factors has a stronger or more significant effect, but it can tell us overall how strong these factors together might be in biasing the evolution of song culture, and if they do indeed change the outcome of song culture compared to the much simpler one-to-one setting.

Methods

Experimental design

We constructed a large isolation chamber from an old 20 cubic ft refrigerator (Fig 4.1). The chamber contained three separate compartments, each equipped with a nesting cups, microphones and video cameras. The chambers were connected, so that bird could easily move between them. Appendix II contains a description of the equipment used in the chambers.

Figure 4.1 | Island Colony setting. Left: an inoperational refrigerator served as a sound chamber to house the colony. Right: middle chamber perch and nest box with colony birds.

We placed an ISO male (a bird that had been raised in complete social and acoustic isolation) and three unrelated females into the sound chamber. They were kept completely isolated, acoustically as well as socially from other birds over a period of 2 years. Once a pair-bond was established, we followed (by audio and video recording) the evolution of this colony. All birds in the colony (except for the 3 female founders) were the descendants of the founder male. However, he only fathered one clutch, after which one of his sons from this first clutch paired up with another of the original female founders and produced all the successive clutches. Based on video observations, we suspect that the rest of the colony birds were all descendents of this pair. Although we did not establish certain genetic relatedness, this pair tended to all the succeeding chicks, hatching them and feeding them until fledging, which has been shown to have a major effect on song tutor selection (Mann & Slater 1994). We allowed the colony to grow with the occasional removal of female offspring over five generations of learners. The colony founder was removed just prior to the hatching of the 5th generation learner, because he was being attacked by the younger males.

Song recording

Song recording in the colony were often contaminated with noise and overlapping singing. To obtain high quality song data, when the young males reached the age of song crystallization (shortly after day 120), the bird was caught and placed in a small sound chamber. After one hour, a colony female was put in the same chamber to induce singing. The two birds remained in the sound box for a couple of hours until at least 10 song bouts were recorded from the males. Afterwards, they were both returned to the colony. In addition to the individual recording session, we recorded and observed the vocalizations produced in the colony.

We used the same sound analysis methods that were described in Chapter 1. We judged imitation at the same three timescales: spectral features, duration of acoustic state and song rhythm. Accuracy was calculated in Sound Analysis Pro 2, using the symmetric and time-courses settings.

Results

4.1 Qualitative assessment of sonograms

Figure 4.2 presents the tentative family tree of our colony males. Clutches are denoted with a letter and a number. The letters indicate families: clutch A was hatched by the founder male, while the B clutches belong to a son from the A1 clutch. The numbers after the letters indicate the successive numbers of clutches belonging to the same parents. As shown, the founder male only reared one clutch, and after sexual maturation, his son assumed the main breeding male role.

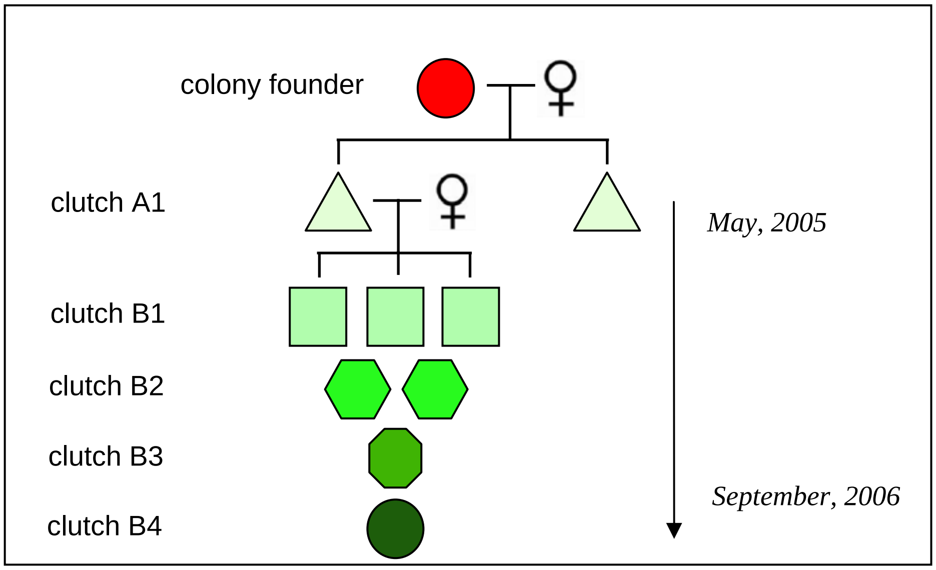
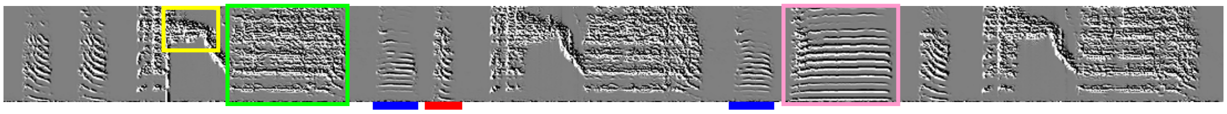


Figure 4.2 | Family tree of island colony. Males are indicated by geometrical symbols, the founder male is the red circle, males belonging to succeeding clutches are indicated by increasingly rounder and darker symbols.

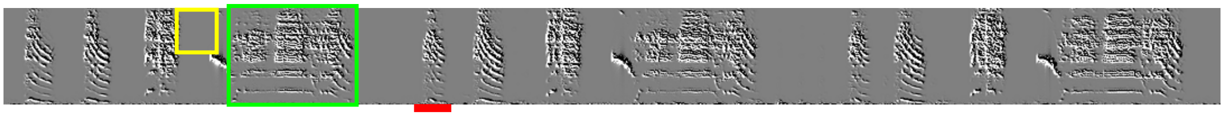
We first judge the development of song culture by looking at the sonograms as we did in the case of WT and ISO birds in Chapter 1. Figure 4.3 presents a sonogram for each of our colony birds. We marked the notes that were especially ISO-like in the founder's song with a yellow, a green and a pink rectangle, and the syllables invented by the pupils were underlined in blue or red. A discussion of the imitations follows below.

Founder of colony (Bird 19)

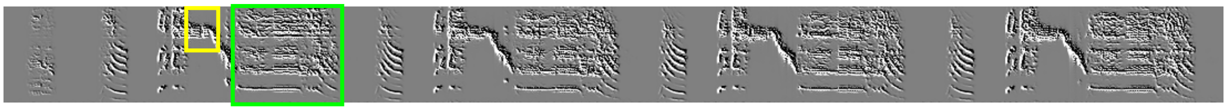
Generation 1 – Clutch A1 (Bird 386)



Generation 1 – Clutch A1 (Bird 1190)



Generation 2 – Clutch B1 (Bird 1147)



Generation 2 – Clutch B1 (Bird 1148)

Generation 2 – Clutch B1 (Bird 1163)

Generation 3 – Clutch B2 (Bird 1193)

Generation 3 – Clutch B2 (Bird 1194)

Generation 4 – Clutch B3 (Bird 1230)

Generation 5 – Clutch B4 (Bird 1254)

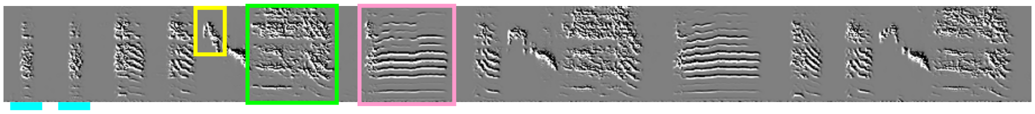


Figure 4.3 | Song sonograms of all colony birds. Top panel: ISO founder's song, descending panels: songs of pupils in succeeding clutches. Colored rectangles highlight ISO-like syllables in founder's song and their imitations in pupils' songs. Syllables improvised by pupils are underlined.

The songs of all the colony males down to five generations contain all or some of the ISO-like syllables of the founder's song. These are the syllables or notes in Fig. 4.3 highlighted with rectangles. All complex notes (yellow and green rectangles) were imitated by all the young birds with 2 exceptions. The first, bird 1190, omitted the yellow note. He was one of the sons of the founder. The other son, bird 386, who fathered all the successive clutches, imitated the yellow note. Since all the rest of the songs contain this note, we can be fairly sure that they mainly imitated their father or their siblings. The other exception was bird 1148 in clutch B1, who only sang call-like syllables. B1 was the only clutch that contained three males, and it is interesting that this abnormal song emerged in this clutch, because as we mentioned at the beginning of this chapter, in a clutch with several male siblings, only 1 or 2 tend to imitate the father's song accurately, and the others often diverge.

The founder of our colony was used in the one-to-one training experiment as well (Tutor 1), and just like in the impoverished condition, his complex and ISO-like syllables were shortened (green syllable) or greatly reduced (yellow syllable). The long call-like syllable (pink rectangle) was somewhat shortened, but not as much as the complex syllables. Only 4 out of 9 birds incorporated the long call into their song motifs.

We observed improvisation in the first and the fifth generations. One of the sons of the founder imitated two new, fairly simple syllables, one of which can be found in the brother's motif, too. When young males are raised in group isolation where they can only hear their peers' vocalizations, they converge on the same syllable types (Volman 1995). Although these brothers were in the presence of their father, the only available song model was an ISO song, which may not be a strong enough stimulus to annul innovation and convergence.

Interestingly, as opposed to the one-to-one tutoring condition, the silence intervals are very short in pupils' songs. This happens by either reducing the silence interval lengths between the repeated motifs as in the song of bird 1147 in Clutch B1 or by filling the silence intervals with short syllables as in the case of bird 1194 in Clutch B2. Both result in a faster and much more WT-like sounding song. Whereas the founder's song sounds very choppy and abnormal, songs of later generation pupils are indistinguishable from WT songs by ear.

Overall, there seems to be a clear progression towards WT song features as in the case of the one-to-one tutoring. We next provide a summary of the imitation of founder syllables by all the colony pupils in section 4.2.

4.2 Similarity measurements of imitations in the colony

Visual assessment of the sonograms revealed to us that in the colony, as well as in the one-to-one tutoring experiment, the pupils changed the tutor song and that the changes accumulated over multiple learning generations. Table 4.1 contains the accuracy measures between the founder's complex syllable and all the succeeding learners' imitations of this syllable. This was the syllable that was most conserved during transmission in the colony. Bird 386 and 1190 (the sons of the founder) produced clearly distinguishable imitations, as mentioned before, because only 386 imitated all of the founder's notes. This bird fathered all the successive clutches, who imitated him or one of their siblings from an earlier clutch. Consequently, as the song is passed down over learning generations, the accuracies steadily decrease (or show no change in clutches B2 and B3). Unfortunately, it is impossible to tell whom the young birds were trying to imitate, but it is interesting that, although the B clutch birds were all siblings, in almost every clutch the progression away from the founder's song continued, so it is safe to say that they took into consideration the songs of the previous clutches not just their father's. Another piece of evidence for this is the copying of syllables improvised by birds in earlier clutches (shown in section 4.1). This is clearly a case of horizontal transmission.

19 (Founder)	386 (A1)	1190 (A1)	1147 (B1)	1148 (B1)	1163 (B1)	1193 (B2)	1194 (B2)	1230 (B3)	1254 (B4)
Accuracy	86%	75%	82%	NA	79%	75%	75%	75%	73%

Table 4.1 | Accuracy comparisons of one syllable between the founder and pupils in succeeding clutches. Decreasing accuracy in succeeding clutches indicates that songs of birds in later clutches are gradually less similar to the founder's song. Note that 386 is the father of the B clutches, so vertically, the decreasing trend holds without exception.

In Table 4.2, we present the imitation statistics for the colony birds. Syllables were subjectively identified and categorized into two groups: complex and call-like. For simplicity, we used the spectral diversity rather than spectral shape as a guideline. In other words, syllables that resemble down-modulated but simple introductory syllables were categorized as call-like, as is the invented, extremely short syllable of bird 1254. None of the invented syllables were substantially modified by imitation, and remained simple in structure. The complex syllable of the founder was imitated by all birds except 1148. For copies of call-like syllables, we gave the identities of the birds who first sang those syllables in parentheses. Only the first and fifth generation birds invented new syllables. The invented call-like syllables of the first generation were copied widely among succeeding clutches and remained prevalent in later generations. Interestingly, in clutches with more than one male (B1 and B2), the siblings always copied different call-like syllables, resulting in divergence. It seems that in the colony, where there was very little song material to imitate and a very low rate of invention, maximum variability between birds was achieved by imitating and modifying the complex syllable and adding call-like syllables from different sources in different combinations. There were no two birds that shared the same sequence of the same syllables.

	19 (founder)	386 (A1)	1190 (A1)	1147 (B1)	1148 (B1)	1163 (B1)	1193 (B2)	1194 (B2)	1230 (B3)	1254 (B4)
Syllable types	3	4	2	1	1	3	2	3	2	3
Complex syllables	1	1	1	1	0	1	1	1	1	1
Call-like syllables	2 (rare)	3	1	0	1	2	1	2	1	1
Copied complex		1	1	1	0	1	1	1	1	1
Invented complex		0	0	0	0	0	0	0	0	0
Copied call-like		1 (19)	0	0	1 (19)	2 (386)	1 (19)	2 (386)	1 (386)	1 (19)
Invented call-like		2	1	0	0	0	0	0	0	1

Table 4.2 | Imitation statistics for colony birds. Syllables are categorized as complex or syllable based on acoustic diversity rather than spectral shape. In the “Copied call-like” row, the bird who was the first one to sing the syllable is given in parentheses. There was no invention of complex syllable. The incorporation of the founder’s introductory note was note taken into consideration, because in most cases it was difficult to tell whether it was a song syllable or an introductory note.

We next examine the shift away from the founder’s song quantitatively by turning to the same three timescales of song structure that we used previously.

4.3 Progression of spectral features towards WT in colony

The sonograms in Figure 4.3 suggest that individual spectral features did not progress steadily in a certain direction over successive clutches in colony songs. The cumulative distribution histograms for spectral features (in Fig. 4.4) suggest the same with the exception of goodness of pitch (Fig 4.4c). Top panels show ISO (red curves) and WT (blue curves) distributions. In the lower panels, the red line indicates the founder bird, and the green lines the birds in successive clutches. For AM (Fig. 4.4a) and FM (Fig.

4.4b) there are no clear differences between the founder and the other colony birds. Goodness of pitch values of colony pupils show a tendency to be below the founder bird's values and this roughly corresponds to the WT distribution, but it is unclear whether there are any progressive differences across successive clutches.

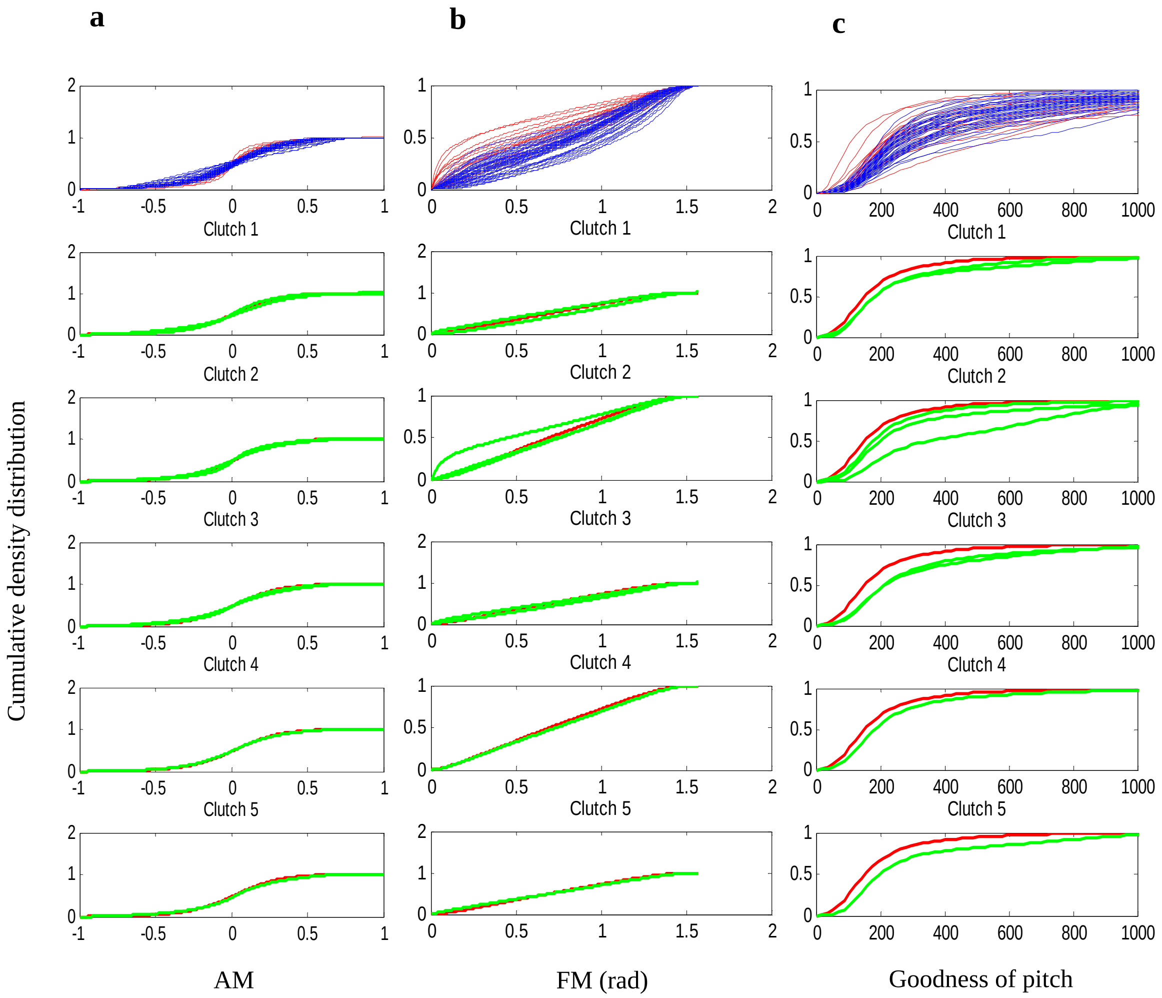


Figure 4.4 | Cumulative density histograms of AM (a), FM (b) and goodness of pitch (c) for colony birds. Top panel: ISO birds (red, n=17) and WT birds (blue, n=52). Lower panels: founder's song (red line) and songs of pupils in succeeding clutches (green lines, $n_{Clutch1}$=2, $n_{Clutch2}$=3, $n_{Clutch3}$=2, $n_{Clutch4}$=1, $n_{Clutch5}$=1).

Individual spectral features did not yield consistent results in the colony, although we observed accumulation of changes in the song sonograms (Fig. 4.3). We now observe these changes in relation to the WT song distribution using PCA of the combined features including FM, AM and goodness of pitch (Fig. 4.5). As shown, all of the colony birds are closer to the WT distribution (in blue) than the founder. Birds of later learning generations (indicated by darker and rounder symbols) are, in general, even closer. Only bird 1148 (clutch B1, light green square) is far away from both the founder and the WT distribution. This was the bird who did not imitate the founder's complex syllable and only sang call-like syllables. His song is extremely unusual and abnormal.

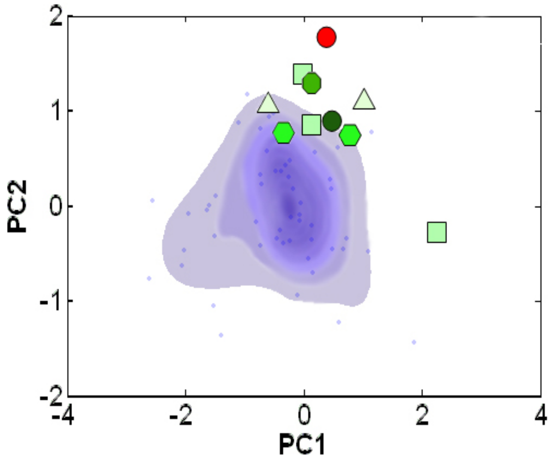


Figure 4.5 | PCA distribution of combined spectral features of colony birds. Darker and rounder symbols indicate successive generations (Fig. 4.2). Blue shape shows the center of WT (n=52) distribution.

4.4 Progression of note-level features in the colony

We have seen that there was a progression towards WT spectral features in later generations of learners in the colony despite the fact that was considerable horizontal transmission. Now we examine note-level features such as the duration of acoustic state and note duration ratios.

A. Duration of acoustic state

As previously, we calculated the duration of acoustic state for the songs of the colony birds. Figure 4.6 shows the probability density of the log of DAS in the founder and in successive clutches. The distribution is very wide in the founder and there are 2 bumps which are preserved and even accentuated in the first and second clutches. After that, though, the distributions become much narrower with one dominant tall peak. This suggests that the variability of DAS is lower in later clutches.

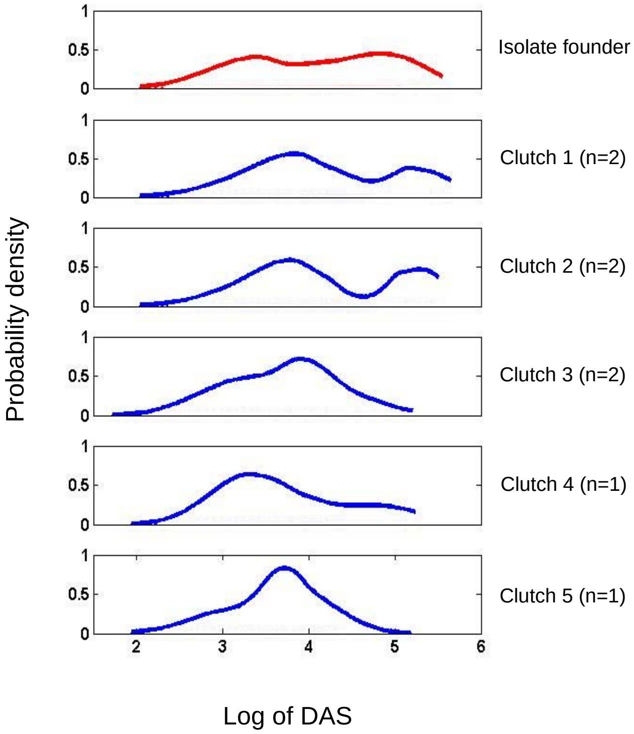


Figure 4.6 | Probability density histograms of log of DAS for colony birds. Top panel: ISO founder's song, descending panels: songs of pupils in succeeding clutches.

In duration of acoustic state, there is a definite direction in the changes in successive clutches. Did these changes make the songs more WT-like? PCA can provide an answer. Figure 4.7 shows the PC1 and PC2 of duration of acoustic state for all colony birds. We can see that with the exception of bird 1148 (light green square in bottom right corner), birds in successive clutches are closer to the WT distribution than the founder. However, the approximation is to a lesser degree than in spectral features, and the fifth generation bird (dark green circle) shifts back towards the ISO founder.

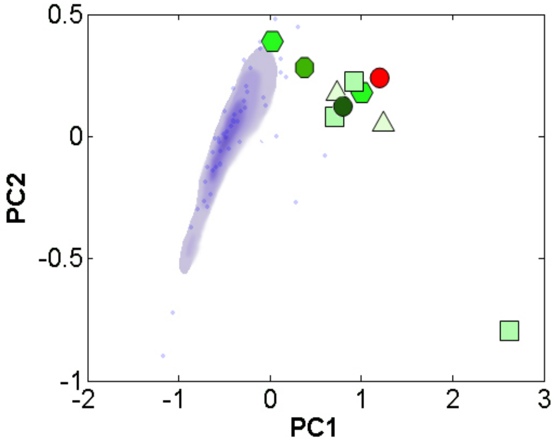


Figure 4.7 | PCA distribution of DAS of colony birds. Darker and rounder symbols indicate successive generations (Fig. 4.2). Blue shape shows the center of WT (n=52) distribution.

As we can see in the sonograms (Fig. 4.3), the song syllables do become shorter over learning generations, although we could not show this as a progression towards WT songs using PCA of DAS (Fig. 4.7). A likely explanation for this is that duration of acoustic state is sensitive to song stability and calls are generally more stable than modulated syllables. In the colony, we observed the prevalence of call-like syllables, one of which was very long, and this could have easily kept overall DAS values near the ISO range. We next examine note length changes using note duration ratios.

B. Note duration ratios

We use note duration ratios to describe the variability of note lengths, high values representing a greater difference between longest and shortest note durations. Since WT notes tend not to be too long, this measure is typically lower in WT birds. Figure 4.8 shows the note duration ratios for complex syllables in all the colony birds. Bird 1148 is omitted from this analysis, because he did not imitate the complex syllable. As we can see, there is a clear decrease in note duration ratios over successive clutches in the colony. This confirms that just as in the one-to-one condition, note duration variability is lower in later generation birds.

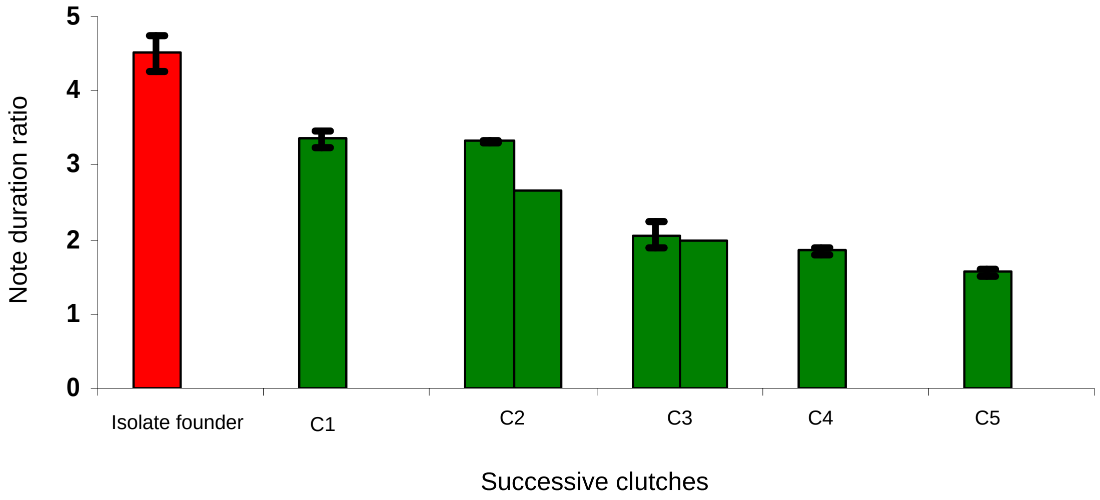


Figure 4.8 | Note duration ratios in songs of colony birds. Red column represents founder, green columns the learners in successive clutches. Error bars indicate within-bird s.e.m.

4.5 Song rhythm progression in the colony

We now turn to the longest timescale song feature, rhythm. After listening to the colony songs, we made an observation that later generation birds sound very much like WT birds. Their rhythms were fast and stable just like in WT song. Indeed, the first two principal components of the rhythm spectra show that the progression towards WT rhythm is robust and consistent over clutches (Fig. 4.9). In nearly every succeeding clutch, the rhythm becomes closer to the center of the WT distribution, and the last couple of birds (darkest symbols) are very near the center. The rhythm PCA of the colony birds confirmed our auditory observations.

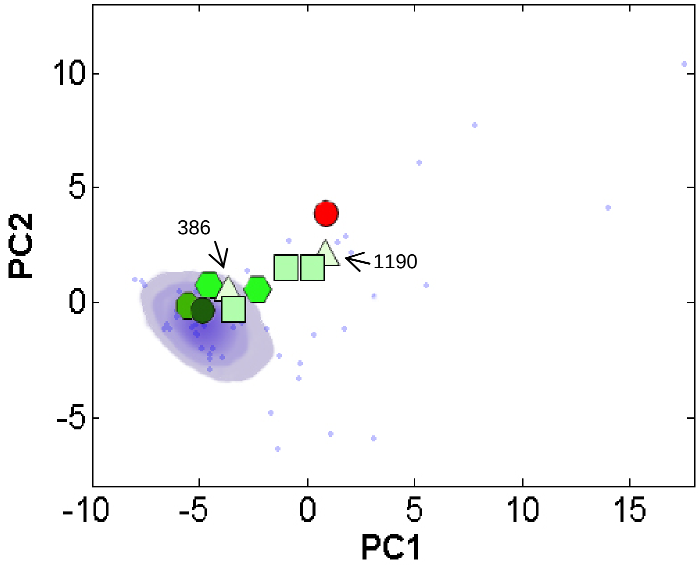


Figure 4.9 | PCA distribution of rhythm in colony birds. Darker and rounder symbols indicate successive generations (Fig. 4.2). Blue shape shows the center of WT (n=52) distribution. The colony founder is represented by the red dot and the two sons of bird 19 (1190 and 386, triangles) are marked with arrows.

4.6 Comparison of imitation between individual tutoring and colony

Since bird 19 participated in one-to-one tutoring as well as the colony experiment, we have an opportunity to compare imitations of his song in the two different conditions. Figure 4.10 shows the imitation of Bird 19's complex syllable over multiple generations and clutches in the one-to-one (Fig. 4.10a) and the colony (Fig. 4.10b) experiment. In both conditions, the long harmonic-like syllable (green rectangle) became shortened, although this happened faster and to a larger degree in the one-to-one condition. The reduction of the broadband, scratchy syllable (yellow rectangle) happened more or less simultaneously in the two conditions. The spectral structure of the green syllable changed much more in the one-to-one condition, because the down-sweep that ended the syllable in Tutor 19's song turned into a very short, non-modulated harmonic (pink arrow) in Generation 2 and became very well defined by Generation 5. This marked the differentiation of a largely homogenous syllable into 2 spectrally different notes. This did not happen in the colony. There, the spectral details of the syllable were preserved (pink arrow), and only the duration decreased.

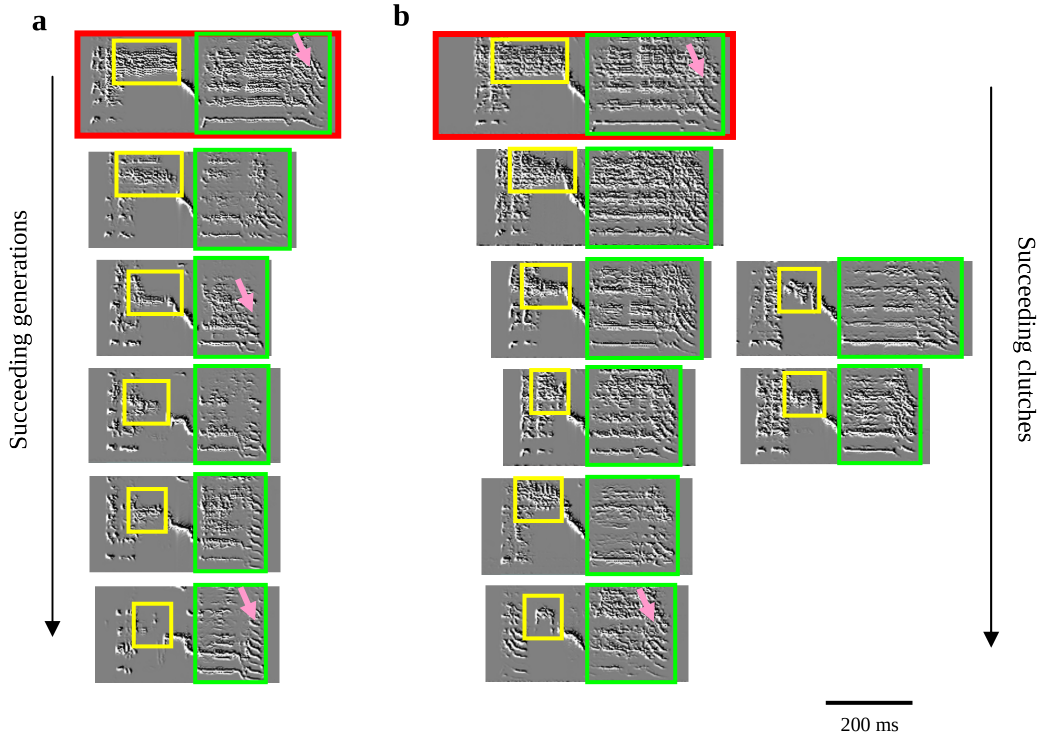


Figure 4.10 | Comparison of imitation in impoverished and rich social conditions. **a.** Imitation of ISO song in one-to-one tutoring experiment. Top panel shows the song of Tutor 19, descending panels are the succeeding generations of pupils. **b.** Imitation of ISO song in colony. Top panel shows the founder (Tutor 19), descending panels are the succeeding clutches. Brothers are shown next to each other. Colored rectangles indicate notes, pink arrow shows syllable differentiation in **a** and the lack of it in **b**.

Comparing the imitations in the two conditions revealed that although the progression towards WT features occurred regardless of the richness of the social environment, there were some important differences between the details of this shift. Spectral changes were more pronounced and more precise in the one-to-one condition, but in the colony birds improvised much more and the songs contained more syllables and were much more rhythmic (Fig. 4.9 and auditory observations). This is not surprising, because zebra finches in a colony interact with each other constantly by calling. These calls can be short and long, and it is very likely that the improvised call-like syllables were derived from contact calls. Juvenile males have been known to incorporate female calls into their songs

(Eales 1985), and in the colony there were many males and females who could have had an influence on the young males.

Discussion

We have presented an experiment that allowed us to study cultural evolution in real-time in a controlled laboratory setting. We investigated what happens to isolate zebra finch song when it is imitated by generations of song learners. We studied this problem in different social settings. We used a recursive training paradigm in an impoverished social setting where young males were housed with adult ISO tutors one-to-one, and upon maturation, they became the tutors of the next generation of learners, and so on. However, since multi-generational changes in a complex social setting may be more representative of natural cultural evolutionary processes, we established a semi-natural colony with an adult ISO male and some females and let them reproduce for a few generations. In the colony young birds had unrestricted interactions with adult males, females, older siblings and clutch mates.

To quantify changes in the song, we developed a method free of qualitative judgment that takes into consideration three levels of song structure: ms-to-ms spectral features, the duration of acoustic state (which describes note-level changes) and song rhythm (which is a characterization of the song at the level of the song bout). We have found that in all song features, regardless of social setting, isolate zebra finch song progresses towards wild-type song over multiple generations of learners, and within 3-4 generations the ISO song features turn into WT features. The extent and time-line of the changes were similar in the two social settings, although spectral features became more structured in one-to-one training and rhythm improved much more in the colony. Since in the colony there was more raw material to imitate (including female calls and siblings' "errors"), there was more song diversity and more variations in song motifs. In the one-to-one setting, birds did not improvise, but instead, often differentiated simple syllables into complex ones.

In the Introduction, we proposed three scenarios that could be responsible for such changes: sexual selection of WT song, selective imitation and biased imitation. Based on our results, we can conclude that the birds used biased imitation to drive the evolution of

WT song from ISO. They imitated ISO-like and WT-like song syllables indiscriminately, but changed the ISO-like features so that in succeeding learning generations those syllables and song features became increasingly WT-like. Since biased imitation was the cause of the changes, females and a complex social environment were not necessary. According to Boyd and Richerson (2005), biased imitation produces nonrandom modifications that can accelerate the rate of evolution. In our experiment, we documented an extremely rapid change in song culture. In the impoverished condition, most of the changes occurred in the first learning generation. In the colony the process was somewhat slower and more gradual. However, in both settings there was a slowing of the progression towards WT in later generations, and the changes seemed to asymptote as the song features approached WT distributions. As noted in the Introduction, this could be because the imitation biases have a smaller effect when the deviation from the WT song culture is smaller, and this deviation was indeed smaller in later generations.

An experiment analogous to ours has recently been conducted in humans (Kirby et al. 2008). Kirby and colleagues trained participants on an “alien” language that included colored objects in motion and random strings of consonant-vowel pairs that were supposed to label these objects. Participants were trained on a set of these labeled moving shapes, which involved looking at the objects and the corresponding labels. After training, participants were tested on a different set that included the set that they were trained on but also objects they had never seen. During testing, participants were provided with a picture and they had to type in the “word” they thought the aliens would produce for that object. Then another group of participants were trained on the output of the last testing session of the previous group. They continued the experiment for 10 generations of learners. They found that transmission error decreased significantly, the largest decrease being in the first 5 generations. Moreover, the increased learnability of the “languages” corresponded to a significant rise in structure. The first presentation of the language was completely random, therefore, it contained no structure at all. However, learners gradually introduced structure by using the same string for the same motion, for example. On occasion, exceptions to rules were reduced by further learning generations.

The outcome of this experiment is analogous to ours in that learners produced increasing structure through an iterated learning process. In both cases, there are innate predispositions to learn something that resembles the WT culture more. Of course, it is fairly clear that in humans random labels that involve rote memorization are more difficult to learn than structured strings with semantic rules. In birds, in the absence of semantics, we cannot be sure that ISO song features are more difficult to learn and that this is what drives the transition to WT, but this could be a plausible hypothesis.

In a sense, our results represent a multigenerational song phenotype that is encoded in every zebra finch but takes multiple generations to appear. It is an extended developmental process. We have observed some similarities between the individual song development in juveniles and the multigenerational development of WT song. For example, in early development, some young males produce back-to-back repetitions of syllable types, but they abandon the repetitions at some point in their development. We saw that syllable repetitions were sometimes present in ISO song, but pupils did not imitate them. It would be interesting to find out if there are other parallels, such as neural and other behavioral changes, that take place during development as well as during the development of the multigenerational WT phenotype.

In the imitation biases, we have identified innate components of culture. Without exposure to song tutoring the biases do not surface to drive the developing song towards WT-like features, but as soon as tutoring is present, the innate biases become activated. It is almost as if they are suppressed in the ISO birds. What is most interesting about our results is that tutoring acts as a “switch” of some sort that turns on the capacity to sing WT-like zebra finch song. We are very interested in the minimum requirement for this capacity. We already saw that simple social interactions are sufficient, but perhaps social interactions are not necessary at all. We are currently training young males with playbacks of ISO songs to investigate this possibility. Moreover, there is a possibility that WT song can be induced within a single bird in the absence of tutoring. When birds are raised in complete social and acoustic isolation, they rely on direct auditory feedback to shape their songs. As a result, they improvise ISO song. However, we know that we do not hear ourselves the same way as others hear us, so if we recorded the developing songs of our isolate birds and played it back to them, we would provide a different level of auditory feedback that may allow the birds to modify their developing ISO songs to be more like WT songs. This experiment is also underway.

We hope that this work, which is currently in press (Fehér et al. 2009), has contributed to the understanding of the development of song cultures, and that it was successful in identifying a good model system to tackle further questions.

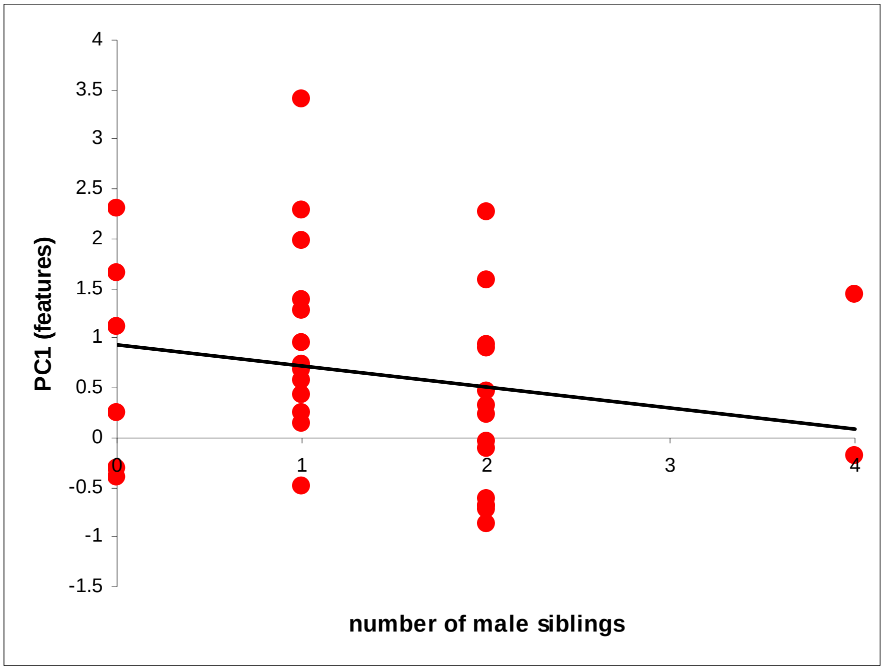
Appendix I. Excluding exposure to adult song in ISO birds

Excluding possible exposure to WT songs prior to day 7

In zebra finches there is no behavioral evidence to early influence of song exposure prior to day 25. Zebra finches are altricial birds: the chicks are helpless and tiny upon hatching, and it takes about 20 days for the auditory system to become fully functional. Measurements obtained on day 10 after hatching show elevated auditory thresholds (about 20-30dB higher than in adults). Therefore, zebra finches do not hear much prior to day 7 post hatch. Zebra finches live in dense social groups and their song is soft and low amplitude, which further decrease the chance of affecting the young chicks. It is also interesting to note that during the first days after hatching the parents tend to be very quiet, and if the males sing at all, they tend to sing away from the nest, facing the other way (Tchernichovski et al. 1998). Of course, the considerations above should not apply to the island colony experiment since chicks hatched there were only exposed to the ISO song.

Excluding possible exposure to ISO subsong prior to day 30

We kept the siblings together until day 30 being well aware that the onset of early subsong is a bit ill-defined. Low-amplitude subsong-like vocalization can be sometimes recorded shortly after fledging using sensitive microphones (we are not sure how common it is, but those sounds are very faint and unstructured). As expected, our analysis show no evidence that the evolution of song culture is affected by the presence of male siblings before day 30. Fig. App. 1 presents the number of male siblings in a clutch, versus the first principal component of the WT/ISO features for all birds used in our study:



Appendix I Figure 1 | Number of male siblings vs. first Principal Component of WT/ISO features of all experimental birds.

As shown, there is no clear trend (r^2=0.04) and looking separately in each experimental groups also show no significant or consistent trend. Similar results were obtained in duration of acoustic state and in rhythm.

Level of isolation in the sound chambers

Our custom made sound chambers provide sufficient level of isolation for zebra finch songs, which rarely exceed 85dB. At 4000Hz (mean frequency of zebra finch songs), we played a 100 dB sound inside the box and measured the sound level outside the box. We measure a sound level of 63 dB, while the baseline noise was 61 dB. Therefore, our boxes contribute a 37 dB sound attenuation. We recorded the entire song development of each bird, and although we listened to recorded sound samples frequently, we heard no traces of songs from other boxes.

Appendix II. Sound chamber equipment for isolated semi-natural colony

Life-sustaining equipment

Jehm Co. linear piston air compressor pumped air through a ¼ inch diameter tube into the left-side chamber. A hole an inch in diameter was drilled in the back of the right-side chamber. This provided sufficient airflow.

The middle chamber was equipped with a 12 inch long 8 Watt light bulb and the 2 outside chambers with 12 inch LED lights. We did this, because the birds voluntarily appointed the middle chamber as a communal feeding area and the outside chambers as nesting areas. The LED lights provided a dimmer light, which we observed had a good effect on nesting.

Water was provided ad libitum in 13 water cups mounted onto the front grid, and food (ABBA seed dry seed and soaked seed mixed with chick food) was placed in small trays in the middle chamber fresh every day.

Monitoring equipment

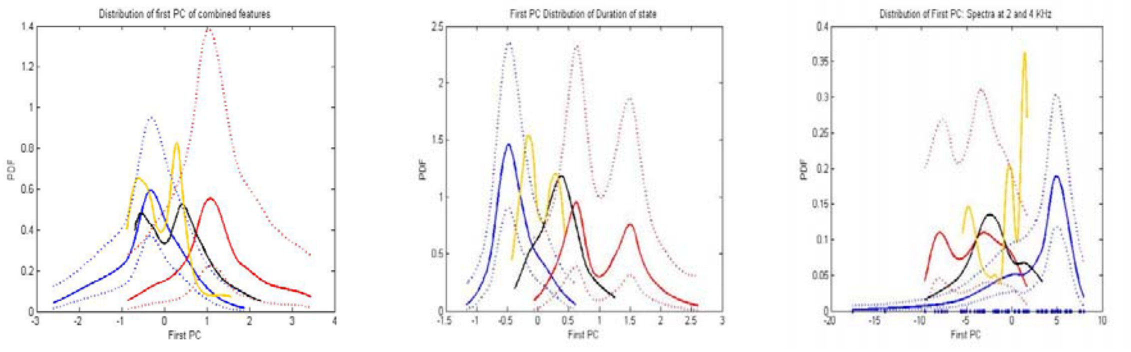
For video monitoring, we used 2 systems in parallel. Labtec webcam USB video cameras were connected to XtraSense surveillance software, and Watec (WAT-230A) and Genwac (GW-230A) color cameras connected to GeoVision surveillance software. The USB cameras gave high-definition colored pictures and simultaneous sound recordings, so for bird identification, this system was much better. The other system was used for reliable continuous observation of nests.

Three microphones were placed in the box, one into each chamber. These recorded sound continuously as long as there were juvenile males (<120 day old) in the colony. Sound recording and analysis was done by Sound Analysis Pro. Developmental song data was studied but not analyzed.

Appendix III. Statistical tests

Distributions of first principal components

In order to assess the effects of recursive tutoring, we considered song feature values of the following four bird groups: WT (52 birds), ISO (17 birds), 1^{st}-generation (13 birds, henceforth denoted as F1), and higher generations (8 birds, denoted as F2+). The distributions of first PC of the four groups are shown in Appendix IV Fig. 1.



Appendix III Figure 1 | Distributions of first PC of combined song features, acoustic state duration, and rhythm spectra. The distributions of ISO, F1, F2+, and WT are shown in red, black, orange, and blue, respectively. The dotted lines are the 95% confidence intervals of WT and ISO. The distinctness between WT and ISO, as well as the shift toward WT due to tutoring, is demonstrated in all three timescales.

The distinctiveness of WT and ISO are very significant. A gradual, yet steady shift of distributions towards WT demonstrates the multi-generational phenotype. The significance of this trend can be subjected to statistical tests.

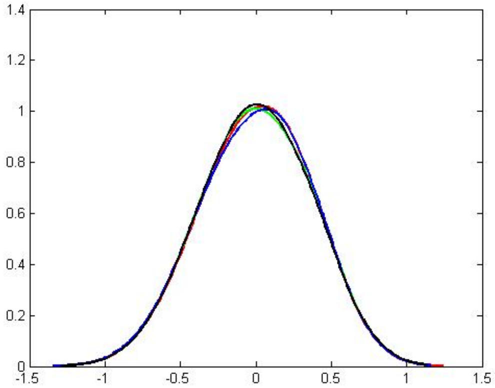
Statistical test methods

The significance of the shift towards WT can be characterized by statistical tests of the difference between the mean values. A two-sided independent *t*-test may be used for this purpose. Since some of the distributions cannot be approximated by Gaussian, we also perform a non-parametric alternative (permutation test) for additional validation. The results from the two methods approximately agree with each other.

Consider two groups of a song trait value (e.g., the first PC of state duration), which have means μ_1 and μ_2, respectively. If the two means are equal under null hypothesis, the division of the whole population into the two groups is arbitrary, as far as mean is concerned. It follows that we can interchange the members of the two groups, without (significantly) changing the following relationship

$$\mu_1 - \mu_2 = 0$$

Interchanging of group members can be realized by permutation. For each (random) permutation, we calculate the statistic $\mu_1 - \mu_2$. We resample through random permutation thousands of times, and obtain a *permutation distribution* of $\mu_1 - \mu_2$. If the null hypothesis holds, the difference between the original group means should not significantly deviate from 0. App. III Fig. 2 shows examples for permutation distributions.



Appendix III Figure 2 | Four examples of permutation distribution for group mean differences of F1 and F2+. They are generated using 10000 permutations and show reasonable stability.

In summary, we test the following null hypothesis:

$$H_0 : \mu_1 - \mu_2 = 0$$

The CDF of the statistic $\mu_1 - \mu_2$ can be obtained from the permutation distribution, and the p-value of the two-sided test is given by

$$p = 1 - [F(|\mu_1 - \mu_2|) - F(-|\mu_1 - \mu_2|)].$$

Six hypotheses are subject to test among the four group means at each time scale. There are 18 hypotheses altogether to be tested on the four birdsong data sets, corresponding to the four bird groups. We apply the *false discovery rate* (FDR) control to correct for multiple comparisons.

Statistical results of group means

<u>i. Spectral frame features</u>

The mean of first PC of the four bird groups are

WT	Higher generations (F2+)	1st-generation (F1)	ISO
-0.4235	0.0266	0.2992	1.2956

The group means clearly demonstrate progression from ISO to WT. The significances of the progressions are statistically tested below.

The p-values using *t*-test are shown below:

	WT	Higher generations	1^{st}-generation	ISO
WT		0.1736	0.0180	< 0.0001
Higher generations			0.4742	0.0045
1^{st}-generation				0.0126
ISO				

The p-values obtained from permutation test are:

	WT	Higher generations	1^{st}-generation	ISO
WT		0.2126	0.0155	< 0.0001
Higher generations			0.4984	0.0104
1^{st}-generation				0.0158
ISO				

The p-value obtained by permutation test may have small fluctuation with different set of 10000 permutations. Hypotheses marked in red can be rejected at the significance level of $\alpha = 0.05$.

Although direct comparison between first generation pupils (F1) and higher generation pupils (F2+) cannot reject the null hypothesis, their relationships to WT and ISO show multi-generational effects:

- F1 is different from WT, while F2+ is not significantly different from WT.
- F1 and F2+ are rather similar, while F1 is very different from ISO. This suggests asymptotic behavior of the recursive tutoring.
- We group higher generations together because there are not enough data of 2^{nd}-generation alone.

ii. Duration of state

The mean of first PC of duration of state are

WT	Higher generations	1^{st}-generation	ISO
-0.3472	0.0195	0.3549	1.0619

Similar observations as spectral frame feature apply.

The p-values obtained from t-test are:

	WT	Higher generations	1^{st}-generation	ISO
WT		0.0089	< 0.0001	< 0.0001
Higher generations			0.0467	< 0.0001
1^{st}-generation				0.0013
ISO				

The p-values obtained from permutation test are

	WT	Higher generations	1st-generation	ISO
WT		0.0069	< 0.0001	< 0.0001
Higher generations			0.0633	< 0.0001
1st-generation				0.0017
ISO				

At 0.05 significance level, the null hypothesis can be rejected for all pairs (t-test). It follows that the four groups are indeed different from each other.

iii. Rhythm

The mean of first PC of different bird groups are

WT	Higher generations	1st-generation (F1)	ISO
-1.98	1.36	2.16	4.09

The p-values obtained from t-test are

	WT	Higher generations	1st-generation	ISO
WT		0.017	0.0033	< 0.0001
Higher generations			0.5872	0.0566
1st-generation				0.1693
ISO				

The p-values obtained from permutation test are

	WT	Higher generations	1st-generation	ISO
WT		0.0775	0.0099	< 0.0001
Higher generations			0.6093	0.075
1st-generation				0.1657
ISO				

At 0.05 significance level, the null hypotheses that can be rejected are marked in red.

iv Multiple comparisons

The 18 hypotheses of comparisons among four groups are tested simultaneously. We use the Benjamini-Hochberg procedure to control the false discovery rate. At FDR level of 0.05, all the hypotheses marked red can be rejected.

Distance metric from WT (tutor-pupil closeness to WT)

In order to study the effects of tutoring, we want to see if the pupil's songs become more similar to the WT songs than the tutor's. Since each bird is represented by a point in an *n*-

dimensional space (e.g., the CDF space of state duration), we can measure the difference between two birds (A and B) by the Euclidean distance between the two vectors:

$$d_{AB}^2 = \sum_{i=1}^{n} [F_A(x_i) - F_B(x_i)]^2 \ .$$

The pupil's song is said to be more similar to the WT songs than the tutor's, if, overall, the distances between the pupil's song and WT songs become smaller than those between the tutor's and WT.

Let us define the tutoring-induced distance change as

$$\Delta_i(P,T) = d_{P,i} - d_{T,i},$$

where $d_{P,i}$ is the distance between pupil's song and the *i*th WT song, and bird *T* is the tutor. The pupil's song is said to become more similar to a given WT song when

$$\Delta_i(P,T) < 0.$$

The probability for the pupil's song to become closer to a given WT song is then given by

$$p_i = P\{\Delta_i < 0\}$$

Since the tutor and its pupil are genetically randomized and never exposed to any WT song, it is reasonable to assume that all the probabilities, p_i, are equal,

$$p_1 = \cdots = p_i = \cdots = p_k = p.$$

We define a random variable *X* to be the number of WT songs for which $\Delta_i < 0$,

$$X = N\{\Delta_i < 0\},$$

which thus follows binomial distribution, $B(n, p)$: each of the Δ_i independently has the same probability *p* to be less than 0. Here, *n* is the total number of WT birds.

Hence, we test the following hypothesis:

$$H_0 : p \le 0.5,$$

$$H_1 : p > 0.5.$$

If the null hypothesis holds, the pupil's songs are no more similar to WT than its tutor's.

The p-values are given in the following table.

Tutoring pair	Feature	State Duration	Rhythm
19 → 1248	< 0.0001	< 0.0001	< 0.0001
19 → 1302	< 0.0001	< 0.0001	< 0.0001
19 → 1340	< 0.0001	< 0.0001	< 0.0001
19 → 1661	< 0.0001	< 0.0001	< 0.0001
1247 → 1315	0.2442	< 0.0001	> 0.9999
1238 → 1342	0.06317	> 0.9999	0.0182
1238 → 1433	< 0.0001	0.0002	< 0.0001
1211 → 1402	< 0.0001	< 0.0001	< 0.0001
1211 → 1566	< 0.0001	0.0039	< 0.0001
1211 → 1655	< 0.0001	< 0.0001	< 0.0001
1249 → 1439	< 0.0001	< 0.0001	< 0.0001
1249 → 1530	< 0.0001	< 0.0001	< 0.0001
1529 → 1622	0.6611	< 0.0001	> 0.9999

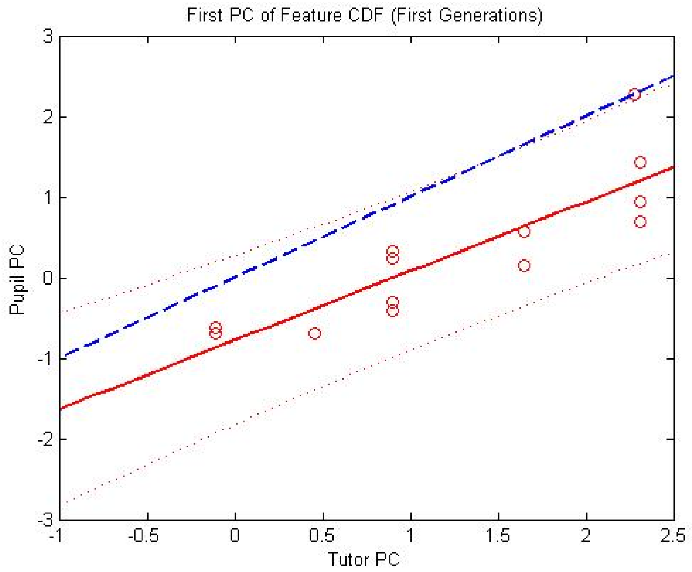
There are 39 hypotheses. Using the Benjamini-Hochberg procedure, those null hypotheses that *cannot* be rejected at a FDR level of 0.01 are marked blue in the table above. We conclude that the pupil's songs become more similar to WT as a result of tutoring.

Tutor-Pupil Correlation: Using First PC of Feature CDF

We now turn to the question how the song trait values of tutor and pupil are related, using the first PC of feature CDF as an example. The relationship at the first generation of tutoring may be approximated by linear regression:

$$y = -0.78 + 0.86x$$

where y is the pupils' PC1, and x represents the tutors' PC. This relationship is shown as the solid red line in Appendix III Fig. 3.



Appendix III Figure 3 | The relationship between tutor's and pupil's trait values. The red solid line is obtained by linear regression, while the dotted lines are the 95% confidence interval. The blue dashed line, $y = x$, represents faithful copying of tutor's song (phenotypic value) by the pupil.

We test if the copying is faithful or partial. Hence the hypotheses are

$$H_0: \quad y = x$$

$$H_1: \quad y = \beta_0 + \beta_1 x, \quad \beta_1 < 1$$

Using the likelihood ratio test and define

$$\lambda = \frac{\max[L(H_0)]}{\max[L(H_1)]}$$

where $L(H_0)$ is the likelihood function under null hypothesis. Under regular conditions, $-2\log(\lambda)$ follows chi-square distribution with 2 degree of freedom. With first PC of feature CDF, we have

$$-2\log(\lambda) = 23.14$$

$$p = 9.4 \times 10^{-6}$$

where p is the p-value. Hence, we shall reject the null hypothesis and conclude that the copying by the pupil is partial.

Increased stability in pupil's song compared to ISO tutor

Pupils' songs were more stable than those of their tutors. Here we examine stability in the duration of syllable renditions (within-type variance). The coefficient of variance (CV) of syllable durations was significantly higher in the isolates' syllables compared to the imitation of those syllables (ISO: CV=15%, range 2-57%; pupils: CV=4%, range 2-9%; p<0.05, Wilcoxon sign test).

BIBLIOGRAPHY

Aisner R. & Terkel J. 1992. Ontogeny of pine cone opening behaviour in the black rat, *Rattus rattus*. Animal Behaviour 44: 327-336.

Baker M. C. & Mewaldt L. R. 1978. Song dialects as barriers to dispersal in white-crowned sparrows, *Zonotrichia leucophrys nuttali*. Evolution 32: 712-722.

Baker M.C. & Cunningham M. A. 1985. The biology of bird-song dialects. Behavioral and Brain Sciences 8: 85-133.

Baker A. J. & Jenkins P. F. 1987. Founder effect and cultural evolution of songs in an isolated population of chaffinches, *Fringilla coelebs*, in the Chatham Islands. Animal Behaviour 35: 1793–1803.

Baptista L. F. 1977. Geographic variation in song and dialects of the Puget Sound white-crowned sparrow. Condor 79:356-370.

Baptista L. F. & Schuchmann K. L. 1990. Song learning in the Anna Hummingbird (*Calypte anna*). Ethology 84: 15-26.

Bonner J. T. 1980. The Evolution of Culture in Animals. Princeton, New Jersey. Princeton University Press.

Boughman J. W. 1998. Vocal learning by greater spear-nosed bats. Proceedings of the Royal Society B 265: 227-233.

Boyd R. & Richerson P. J. 2005. The Origin and Evolution of Cultures (Evolution and Cognition). Oxford University Press, USA.

Campbell D. T. 1965. Variation and selective retention in socio-cultural evolution. In: Social change in developing areas: A reinterpretation of evolutionary theory, eds. Barringer H. R., Blanksten G. I & Mack R. W. Schenkman, Cambridge, Mass. 19-49.

Cavalli-Sforza L. L. & Feldman M. W. 1981. Cultural transmission and evolution: a quantitative approach. Princeton University Press, Princeton, N. J.

Chilton G. & Lein M. R. 1996. Long-term changes in songs and song dialect boundaries of Puget Sound white-crowned sparrows. Condor 98: 567-580.

Clayton N. S. 1987. Song tutor choice in zebra finches. Animal Behaviour 35: 714-721.

Darwin C. 1871. The Descent of Man, and Selection in Relation to Sex. John Murray, London.

Dawkins R. 1976. The Selfish Gene. Oxford University Press.

Doupe A. J. & Kuhl P. K. 1999. Birdsong and human speech: common themes and mechanisms. Annual review of neuroscience 22: 567-631.

Du P. & Troyer T. W. 2006. A segmentation algorithm for zebra finch song at the note level. Neurocomputing 69(10-12): 1375-1379.

Goodall J. 1964. Tool-using and aimed throwing in a community of free-living chimpanzees. Nature 201: 1264-1266.

Eales L. A. 1985. Song learning in zebra finches: some effects of song model availability on what is learnt and when. Animal Behaviour 33: 1293-1300.

Fehér O., Wang H., Saar S., Mitra P. P. & Tchernichovski O. 2009. De novo establishment of wild-type song culture in the zebra finch. Nature, published online before print May 3, doi: 10.1038/nature07994.

Foote A. D., Griffin R. M., Howitt D., Larsson L., Miller P. J. O. & Hoelzel A. R. 2006. Killer whales are capable of vocal learning. Biology Letters 2(4): 509-512.

Gardner T. J., Naef F. & Nottebohm F. 2005. Freedom and rules: the acquisition and reprogramming of a bird's learned song. Science 308: 1046-1049.

Henrich J. & Boyd R. 2002. Culture and cognition: why cultural evolution does not require replication of representations. Culture and Cognition 2: 87–112.

Hunt G. R. 1996. Manufacture and use of hook-tools by New Caledonian crows. Nature 379: 249-251.

Hunt G. R. and Gray R. D. 2003. Diversification and cumulative evolution in tool manufacture by New Caledonian crows. Proceedings of the Royal Society, London B 270: 867-874.

Immelmann K. 1969. In: Bird Vocalizations: Their Relations to Current Problems in Biology and Psychology, ed. Hinde R. A. Cambridge University Press, Cambridge. 61-74.

Kegl J. 2002. Language Emergence in a Language-Ready Brain: Acquisition Issues. In: Language Acquisition in Signed Languages, eds. Morgan G. & Woll B. Cambridge University Press. 207-254.

Konishi M. 1965. The role of auditory feedback in the control of vocalization in the white-crowned sparrows. Z. Tierpshychol. 22: 770-783.

Leitner S., Nicholson J., Leisler B., DeVoogd T. J. & Catchpole C. K. 2002. Song and the song control pathway in the brain can develop independently of exposure to song in the sedge warbler. Proc. Biol. Sci. 269: 2519-2524.

Mann N. I & Slater J. B. 1994. Song tutor choice by zebra finches in aviaries. Animal Behaviour 49: 811-820.

Marler P. & Tamura M. 1962. Song "dialects" in three populations of white-crowned sparrows. Condor 64: 368-377.

Marler P. & Tamura M. 1964. Song 'dialects' in three populations of white-crowned sparrows. Science 146: 1483-6.

Nelson D.A. & Marler P. 1994. Selection-based learning in bird song development. Proceedings of the National Academy of Sciences, USA 91: 1498-1501.

Noad M. J., Cato D. H., Bryden M. M., Jenner M.-N. & Jenner K. C. S. 2000. Cultural revolution in whale songs. Nature 408: 537.

Nottebohm F. 1966. The role of sensory feedback in the development of avian vocalizations. Ph.D. Thesis, University of California, Berkeley.

Nottebohm F. & Arnold A. P. 1976. Sexual dimorphism in vocal control areas of the songbird brain. Science 194: 211-213.

Nottebohm F., Nottebohm M. E., Crane L. A. & Wingfield J. C. 1987. Seasonal changes in gonadal hormone levels of adult male canaries and their relation to song. Behavioral and Neural Biology 47: 197-211.

Okanoya K. 2004. The Bengalese Finch: A Window on the Behavioral Neurobiology of Birdsong Syntax. In: Annals of the New York Academy of Sciences, Behavioral Neurobiology of Birdsong. 1016: 724-735.

Payne R. B. 1981. Song learning and social interaction in Indigo Buntings. Animal Behaviour 29: 688-697.

Payne R. B. 1985. Behavioral continuity and change in local song populations of Village Indigobirds *Vidua chalybeata*. Z. Tierpsychol. 70: 1-44.

Pepperberg I. M., Gardiner L. I. & Luttrell L.J. 1999. Limited contextual vocal learning in the grey parrot (*Psittacus erithacus*): the effect of co-viewers on videotaped instruction. Journal of Comparative Psychology 113: 158-172.

Price P. H. 1979. Developmental determinants of structure in zebra finch song. Journal of Comparative and Physiological Psychology 93: 260.

Range F., Viranyi Z. & Huber L. 2007. Selective Imitation in Domestic Dogs. Current Biology 17(10): 868-872.

Reiss D. & McCowan B. 1993. Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): evidence for vocal learning. Journal of Comparative Psychology 107(3): 301 - 312.

Rose G. J., Goller F., Gritton H. J., Plamondon S. L., Baugh A. T. & Cooper B. G. 2004. Species-typical songs in white-crowned sparrows tutored with only phrase pairs. Nature 432: 753-758.

Saar S. & Mitra P. P. 2008. A technique for characterizing the development of rhythms in bird song. PLoS ONE 3(1): e1461. doi: 10.1371/journal.pone.0001461.

Sanz C., Call J. & Morgan, D. 2009. Design complexity in termite-fishing tools of chimpanzees (*Pan troglodytes*). Biology Letters. Published online before print March 4, 2009, doi: 10.1098/rsbl.2008.0786.

Searcy W. A., Marler P. & Peters S. S. 1985. Songs of isolation-reared sparrows function in communication, but are significantly less effective than learned songs. Behavioral Ecology and Sociobiology 17: 223-229.

Slabberkoorn H. & Smith T. B. 2002b. Habitat-dependent song divergence in the little greenbul: an analysis of environmental selection pressures on acoustic signals. Evolution 56(9): 849-1858.

Smith V. A., King A. P. & West M. J. 2000. A role of her own: female cowbirds, *Molothrus ater*, influence the development and outcome of song learning. Animal Behaviour 60: 599-609.

Tchernichovski O., Schwabl H. & Nottebohm F. 1998. Context determines the sex appeal of male zebra finch song. Animal Behaviour 55: 1003-1010.

Tchernichovski O., Nottebohm F., Ho C. E., Bijan P. & Mitra P. P. 2000. A procedure for an automated measurement of song similarity. Animal Behaviour 59: 1167-1176.

Thielcke G. Geographic variation in bird vocalizations. 1969. In: Bird Vocalizations: Their Relations to Current Problems in Biology and Psychology, ed. Hinde R. A. Cambrindge Univ. Press, Cambridge. 311-339.

Thompson A. D. Jr. & Baker M. C. 1993. Song dialect recognition by male white-crowned sparrows: effects of manipulated song components. Condor 95: 414-421.

Thorpe W. H. 1958. The learning of song patterns by birds, with especial reference to the song of the chaffinch *Fringilla coelebs*. Ibis 100: 535–570.

Tierney A. T., Russo F. A. & Patel A. D. 2008. Empirical comparisons of pitch patterns in music, speech, and birdsong. *Paper presented at Acoustics '08 (155th meeting of the Acoustical Society of America), July 2008, Paris, France.*

Trainer J. M. 1983. Changes in song dialect distributions and microgeographic variation in song of white-crowned sparrows (*Zonotrichia leucophrys muttalli*). Auk 100: 568-582.

Volman S. F. & Khanna H. 1995. Convergence of untutored song in group-reared zebra finches (*Taeniopygia guttata*). Journal of Comparative Psychology 109(3): 211-221.

West M. J., King A. P. & Eastzer D. H. 1981. Validating the female bioassay of cowbird song: relating differences in song potency to mating success. Animal Behaviour 29: 490–501.

Williams H. 1990. Models for song learning in the zebra finch: fathers or others? Animal Behaviour 39: 745-757.

Williams H., Kilander K. & Sotanski M. L.1993. Untutored song, reproductive success and song learning. Animal Behaviour 45: 695-705.

)