Chapter to appear in 'The Evolutionary Emergence of Language' Oxford University Press

FoxP2 and deep homology in the evolution of birdsong and human language

Alan Langus[1,2], Jana Petri[3,] Marina Nespor[1,2], Constance Scharff[3]

[1]Università degli Studi di Milano
Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy

[2]Scuola Internazionale Superiord di Studi Avanzati SISSA
via Bonomea, 265, 34136 Trieste TS, Italy

[3]Freie Universität Berlin, Takustr. 6 14195 Berlin, Germany

Corresponding authors:
Constance Scharff scharff@zedat.fu-berlin.de
Marina Nespor marina.nespor@unimib.it
Alan Langus alanlangus@gmail.com
Jana Petri petri@molgen.mpg.de

Abstract

Spoken language and birdsong share a number of striking parallels. We summarize the biologically tractable cognitive abilities necessary for spoken language and for birdsong and argue that the similarities are not limited to sensorimotor processes – but may extend to the conceptual and computational systems. We survey converging evidence for the relevance of the *FoxP2* gene and its associated molecular network for language, and its role in the acquisition and production of birdsong. Many questions regarding the similarities between spoken language and birdsong remain unanswered, but increasing evidence suggests that human and non-human communication systems may rely on conserved molecular toolkits that act as genetic modules. These may specify the neural circuits subserving these particular behaviors, and organize their function. Elucidating these genetic modules in different animal models promises insights into the evolution of language and other complex traits.

1. Introduction

Human language constitutes a fascinating evolutionary puzzle because non-human animals are usually considered not to have language. This raises important questions as to how, when, and why this trait evolved in the human lineage. Language in its entirety comprises the sensorimotor system that deals with the perception and production of speech sounds or gestures, and the meaning system that deals with reasoning and inference. Spoken language, e.g. ‘speech’, refers to the sensorimotor aspects of language. Speech and birdsong share many features and therefore constitute an interesting source of information about the evolution of human language. For example, both rely on structurally complex communicative vocalizations that require exposure to and imitation of adult vocalization, extensive practice, at the same time being constrained by innate predispositions (for an overview see Bolhuis et al., 2010). There are similarities between the ways the sensory experience of species-specific vocalizations in humans and birds is internalized and used to shape vocal outputs, raising the possibility that spoken language and birdsong may depend on similar neural mechanisms (Doupe & Kuhl, 1999). Converging evidence suggests that birdsong and spoken language may also be governed by homologous genetic mechanisms that are conserved across species (Fisher & Scharff, 2009; Scharff & Petri, 2011). These similarities are particularly striking considering that birdsong and speech arose in distinct evolutionary branches that diverged some 300 million years ago (Kumar & Hedges, 1998), suggesting that certain aspects of vocal communication in humans and birds may constitute a case of deep homology (Carroll, 2005).

The inclusion of concepts and methods from developmental biology in evolutionary theory has led to the field of evolutionary developmental biology, or 'evo-devo' (Gilbert et al, 1996). Drawing on molecular and genetic methods, developmental biologists have uncovered conserved molecular networks that shape the morphology of different species (Stern & Orgogozo, 2009) that may not only apply to the evolution of form, but extend also to the evolution of behavior (Robinson et al., 2008; Bertossa, 2011). For example, the FoxP2 transcription factor and its regulatory network with which it is interacting may be part of a molecular toolkit that is essential for sensory guided motor learning in cortico-striatal and cortico-cerebellar circuits, in e.g. mice, humans, and songbirds (Fisher & Scharff, 2009; Scharff & Petri, 2011). This transcription factor and its associated molecular network may thus constitute one of the constraints that channel evolutionary patterns towards similar outcomes, e.g. learned vocal communication in diverse taxa. This evo-devo framework can thus provide a useful theoretical framework to the study of 'genetic modules' that are necessary for human language.

2. Biologically tractable units of language and birdsong

It is possible to identify and entangle different modular cognitive systems responsible for specific linguistic tasks in humans and other species. In broadest terms, it is possible to identify distinct modular cognitive systems responsible for specific linguistic tasks (Chomsky, 1957; Fodor, 1983). There is good agreement across most linguistic theories that the production and comprehension of language require at least three task-specific cognitive systems: the conceptual-intentional system (semantics) that provides and interprets the meaning of linguistic utterances; the sensory-motor system (phonology and phonetics) that produces and perceives the actual sounds and signs of language; and the computational system of grammar (syntax) that links meaning with sounds by generating the structure of sentences (Hauser et al., 2002; Pinker & Jackendoff, 2005).

It is generally believed that birdsong shares parallels with speech (the sensory-motor component of spoken language) but less attention has been paid to the possibility that homologues exist also at the level of the conceptual and computational systems. The molecular mechanisms that change morphological form and behavioral norms during evolution function in a graded manner. Mutations can modify gene products, altering their function. Alternatively, the place, time or amount of a particular gene product can change because of mutations in regulatory regions. Therefore cross-species behavioral differences are also not likely to be categorical but graded. We favor the view that different attributes of languages exist in principle in other species, to varying degrees and with potentially different consequences (Thorpe, 1974; Zirin, 1980) rather than categorical statements such as 'language has unique attributes that do not exist in animals'. Thus even though birdsong is by no means identical to language, birds may still utilize cognitive mechanisms and neural substrates for song similar to those that humans use for language. This raises the question of whether the three cognitive systems necessary for language can be identified also in birdsong. This review summarizes vocal communication only, but we do not exclude that parallels between human sign language and animal gestures exist, because human sign language and vocal language also share certain features (ranging from mechanisms of acquisition to phonological, syntactic and semantic structures) (Klima & Bellugi, 1979).

(a) Sensory-motor processes

Birdsong and human speech are complex acoustic signals, both produced by controlling the airflow in the vocal system. While the structure of the vocal tracts of humans (Stevens, 1994) and songbirds (Gaunt 1987, Goller & Larsen 1997, Suthers 1997, Fee et al 1998, Elemans et al 2010) differ, both birdsong and speech production require specialized neural and physiological control, as well as coordination between respiration and a number of vocal motor-programs (Doupe & Kuhl, 1999). While the brain regions required for speech perception are spread from the auditory cortex to the left temporal and frontal cortices, the regions responsible for motor-programming and co-coordinating speech production are traditionally thought to involve the inferior frontal cortex (including Broca's area). Songbirds also have specialized, discrete brain regions controlling song production and perception, the song motor pathway (SMP), including an ascending auditory pathway, a posterior motor pathway connecting the cortical/pallial song control region HVC via a number of relay stations to the motor neurons controlling the muscles of the sound source, the syrinx, and a third, anterior pathway essential for acquisition of song and for adult song plasticity (for an overview see Bolhuis et al., 2010).

In human language, phonemes are combined to form syllables that constitute the basic units of speech. Likewise, in birdsong individual acoustic elements (frequently called notes) are often grouped together into multi-element units (often called syllables), the basic processing units of birdsong (Cynx, 1990). In both humans and songbirds, the sensitivity to syllables already exists in infancy (Bertoncini et al., 1988; Nelson & Marler, 1993; Perani et al., 2011). However, in contrast to spoken language, where the different combinations of individual phonemes lend human language much of its open-ended 'unbounded' power of expression, in all investigated bird species, there are more song elements in the repertoire than combinations of those elements (Hultsch & Todt, 2001), which has led to the assumption that birdsong cannot be unbounded (but see below). Syllables, and groups of hierarchically ordered syllables (called phrases or motifs in birdsong) are separated by silent intervals that may aid perceptual grouping of song elements (Anderson, Dave & Margoliash, 1996), just like prosodic cues such as final lengthening and pitch declination signal phonological and intonational phrase boundaries in speech (Lehiste, 1970; Selkirk, 1984; Nespor & Vogel, 1986; for an overview see Cutler et al., 1997 and Langus et al. under revision).

Finally, in humans, the perception of speech is at least partially segregated from other auditory events and sub-served by speech perception circuits in the superior temporal cortex (Wernicke's area and surrounding regions) (Hickok & Poeppel, 2000; Poeppel, 2001; Trout, 2001; Vouloumanos et al., 2001) some of which appear to be functional already in newborn infants (Peña et al., 2003; Perani et al., 2011). The superior temporal cortex is active during speech perception in neonates (Imada et al., 2006) and Broca's area only engages in speech perception when infants start babbling (Dehaene-Lambertz et al., 2006; Imada et al, 2006), suggesting a role for both innate as well as developmental factors in the development of the neural substrates for speech perception. However, the representation of speech must be specific enough for newborn infants to discriminate phonemes categorically (Eimas et al., 1971), perceive well-formed syllables as units (Bertoncini & Mehler, 1981; Bertoncini et al., 1995; Bijeljac-Babic et al., 1993), and distinguish languages on the basis of rhythmic classes (Nazzi et al., 1998). In fact, as infants fail to distinguish languages from different rhythmic classes when speech is played backwards (Mehler et al., 1988), representations seem to be speech specific.

The fact that songbirds require exposure to adult song for song acquisition (Marler 1970, Thorpe 1958), and show preferences for con-specific song in infancy (Nelson & Marler, 1993) and adulthood (Dooling & Searcy 1980), suggests that the sensory-motor processes and representations involved in song perception and production may be geared towards species-specific vocalizations in songbirds just like in humans (Doupe & Kuhl, 1999). Importantly, specialized, discrete brain regions have been identified for song recognition (called first and second auditory regions, including the caudomedial nidopallium (NCM) and caudo medial mesopallium (CMM)) (Mello et al., 1992 Bolhuis et al., 2010). Some neurons in the song control system show selective responses to conspecific song (Margoliash, 1983; Williams & Nottebohm, 1985; Margoliash, 1986; Doupe & Konishi, 1991), others for birds' own song (Solis et al., 2000; Margoliash & Konishi, 1985), and only a small proportion respond to the tutor song (Solis et al., 2000). Interestingly, the neurons encoding and decoding birdsong show strong responses to bird's own song, but no response when the song is played backwards (Hahnloser & Kotowicz, 2010; Doupe & Konishi, 1991). In summary, both spoken language and birdsong rely on sensory-motor representations and processes that are specialized for producing and perceiving conspecific vocalizations.

However, the extent to which the sensory-motor systems in humans and songbirds are homologous has yet to be delineated. For example, in speech the sensory-motor processes contribute to various aspects of language. Sensory-motor systems are involved in generating and interpreting the different phonemic combinations of our spoken utterances. By manipulating prosodic cues such as pitch, duration and intensity of our utterances we can convey many different things. These cues can provide information about syntax (Nespor et al, 2008; Bion et al 2011), emotional states (e.g. irony or sarcasm) (Mozziconacci, 2002), the type of statement, e.g. questions (Lieberman, 1967; Bolinger, 1989), emphasis (Selkirk, 1995; Ladd, 1996), or about the meaning of words. For instance, vowel length is phonemic in Estonian and can be used to differentiate lexical entries, e.g. ma „I", maa „land"; pitch is similarly phonemic in tonal languages like Chinese) (Lehiste, 1966; San, 2007). Are there equivalents of prosodic clues in birdsong and if so, which ones? How can these questions be addressed?

There is some evidence that birds manipulate the properties of their song to convey affective information in a way that resembles humans' expression of emotions with prosody. Many songbirds sing songs in different social contexts e.g. courtship or territorial defense. Song characteristics, e.g. stereotypy, speed, and number of elements can vary between these contexts (Sossinka and Böhner, 1980 Kroodsma & Byers, 1991). Furthermore, during vocal practice juvenile zebra finches can already switch to courtship song that is strikingly more stereotyped than juvenile 'plastic' song (Kojima & Doupe, 2011), suggesting that the social context of birdsong can influence the quality of the song already from a very early age.

In spoken language the prosody of the speech signal contains information about syntactic structure by automatically organizing the units marked by suprasegmentals – pitch, duration and intensity – in the prosodic hierarchy that at least partially mimics the hierarchical structure of syntax (Selkirk, 1984; Nespor & Vogel, 1986; Beckman & Pierrehumbert, 1986; Hayes, 1989). The prosodic constituents most relevant for signaling syntactic structure are the phonological phrase and the intonational phrase. The phonological phrase extends from the left edge of a phrase to the right edge of its head[1] in head-complement languages; and from the left edge of a head to the left edge of its phrase in complement-head languages (Nespor & Vogel 1986). The constituent that immediately dominates the phonological phrase is the intonational phrase that is coextensive with intonation contours, thus accounting for natural break points in speech (Pierrehumbert & Hirschberg, 1990). Prosodic units at one level of the prosodic hierarchy are always exhaustively contained in the prosodic units that dominate them. Whether suprasegmental cues exist in birdsong and whether they signal the structure of song has not been studied.

Finally, is the perception of song in birds guided by acoustic biases similar to those that humans have for speech perception? For example, the human auditory system groups elements according to the Iambic-Trochaic Law (ITL) – i.e. elements that alternate mainly in duration are grouped iambically (weak-strong) and elements that alternate mainly in intensity are grouped trochaically (strong-weak) (Hayes, 1995). Interestingly, because head-final languages (e.g. Turkish and Japanese) mark prominence in phonological phrases initially through pitch and intensity and head-initial languages (e.g. Italian and English) mark prominence in phonological phrases finally through duration (Nespor et al., 2008), the ITL may even guide infants' discovery of word order relations during language acquisition (Bion et al., 2011). Acoustic biases like the ITL, that operate over both linguistic and non-linguistic sound-sequences (Hay & Diehl, 2007), could thus enable us to determine the extent to which characteristics of the sensory-motor system in humans are species-specific or may be a result of general principles also existing in other species such as songbirds.

(b) Conceptual-intentional processes

Spoken language can express arbitrary thoughts with open-ended semantics, which is a feature usually considered to be exclusive to humans (Hockett, 1960). Researchers are operating under the assumption that animals communicate mainly about 'fighting' or 'flirting' in a non-compositional way using semantically holistic sounds (Bickerton, 1990; Hurford, 2008; Fitch, 2009), even though there are not enough data to exclude that animal communication is more semantically loaded than currently thought (Scharff & Petri, 2011). For example, bees communicate 'about' food (Frisch, 1967) and many animals use specific calls that refer to different predators (Rendall, et al., 2009; Seyfarth et al., 1980; Slobodchikoff et al., 2009). While these systems certainly do not reach the complexity of the semantics of human language, they indicate that animals are in principle able to use arbitrary vocal or other gestures associated with meaning. Some core characteristics of conceptual-intentional processes have been observed in birds. For example, just like human language depends on the behavioral contexts such as social communication vs. 'inner speech', birds also communicate in many different contexts (Kroodsma & Byers, 1991) that include mate attraction (Langmore, 1998) and parent-offspring interactions (Slater et al., 1988) and predator alarms (Griesser, 2008). In addition, birds may even vocalize outside of any obvious communicative context, as is the case when male zebra finches sing 'undirected song' which often occurs while birds are alone (Sossinka & Böhner, 1980). Furthermore, evidence that is consistent with an interpretation of intentional manipulation of vocal behavior exists in western scrub-jays that appear to conceal auditory information when competitors can hear them but cannot see them (Stulp et al., 2009). Finally, also 'displacement' – the ability to refer to absent events, things or concepts, often considered to be restricted to humans (Hockett, 1960; Bickerton, 2010) – may in principle extend to birds as evidence from the corvid family shows that birds may act with an eye to future events (Emery & Clayton, 2004). Taken together, these findings raise the possibility that conceptual-intentional processes may turn out to be more common in animals than currently thought.

It is important to note that there are few experiments that have investigated 'the meaning of parts' in birdsong and consequently there is no evidence for 'productivity', the ability to create new utterances by combining existing utterances (Hockett, 1960). For example, in human language words can be formed through agglutination (adding affixes: e.g. shame + less = shameless), blending (joining parts of several words: e.g. *smoke* and *fog* to form *smog*), compounding (stringing together

---

[1] The syntax of natural languages consists of clauses that are formed of phrases. Syntactic phrases consist of a head and optionally one or more complements. The head of a phrase is the word that determines the syntactic type of the phrase of which it is a member. In the sentence *some dogs chase kids*, *chase* is the head of the verb phrase *chase dogs*. In a head-initial language (e.g. English and Italian) the heads of phrases precede their associated complements: for example, verbs precede their objects, prepositions precede nouns and main clauses precede subordinate clauses. In addition, specifiers – syntactic categories that specify the heads, as for instance 'some' in some dogs' precede the head they are associated with. In contrast, in head-final languages (e.g. Turkish and Japanese) the complements precede their heads.

words: e.g. dark and room to form darkroom), conversion (forming a new word from an existing one of a different category: e.g. verbification of nouns such as dress) (Bauer, 1983). The underlying processes can be rule-governed processes (e.g. past tense of regular verbs: fax - faxed) as well as associative processes (e.g. past tense of irregular verbs: go – went) (Pinker, 1991). To our knowledge, few studies have addressed the question whether different combinations of calls or song elements could have different meanings. However, it is also possible that the 'combinatorial units' in birdsong are not the smallest song elements (i.e. notes and syllables – the equivalents of phonemes and syllables in speech) but the song elements that occupy higher levels in the structural hierarchy of song, e.g. a string of ordered elements sung together (called motif or phrase, or song type in the literature). For example, in some species with large repertoires, song types occur within long, non-random, and non-unidirectional arranged sequences. Such song bouts can last from minutes to even many hours (Kipper et al., 2004; Hultsch & Todt, 1989; Kroodsma & Parker, 1977; Suzuki et al., 2006). At this organizational level, there is a lot of room for complex sequential rules that could be sophisticated enough to carry semantic information. In fact, motif order in many birds and some bats (Bohn et al., 2009) can be much more dynamic than note order within a motif. Thus, while the evidence suggests that intentional processes and preliminary conceptual capacities do exist in birds, their scope has yet to be determined.

(c) Computational system

The precise nature of the computational system necessary for language is not fully understood. Some theories postulate that the grammatical diversity observed among the world languages is genetically encoded in the human computational system (Chomsky, 1980; Pinker & Jackendoff, 2005). This implies that representations of all the rules of world's languages are fully predetermined in the computational system and the infant learner converges through positive evidence from linguistic input during language acquisition on a specific grammar. On the other hand, recent approaches in linguistics assume that the computational system only generates the structure of the world's languages with a limited number of basic rules (e.g. *merge* and *move*) (Chomsky, 1995; Hauser et al., 2002). According to such a view, the specific grammatical configurations (e.g. word order) are the byproducts of the interaction between the three cognitive systems necessary for human language (Chomsky, 1995; Langus & Nespor, 2010). However, regardless of the specific points of view, in spoken language, the computational system of grammar links meaning with sound (Hauser et al., 2002; Pinker & Jackendoff, 2005).

Both birdsong and human language are hierarchically organized according to syntactic constraints (Bolhuis et al., 2010; Berwick et al. 2011). In birdsong, individual notes can be combined into syllables, syllables into 'motifs', and motifs into a complete song. The song structure depends on the particular ordering regularities specific to a given species (Marler & Slabbekoorn, 2004). Similarly to spoken language, birdsong thus arranges discrete acoustic elements in specific temporal orders (Okanoya, 2004; Sasahara & Ikegami, 2007). These may be relatively fixed sequences with little variation (e.g. zebra finches, chaffinches or white-crowned sparrow), or show more systematic variation (e.g. Bengalese finches, starlings, winter wrens, or nightingales, where one element might be followed by several others, with overall song structure governed by probabilistic rules between a finite number of states (Catchpole & Slater, 2008; Wohlgemuth et al. 2010)).

The syntax of birdsong is usually referred to as phonological syntax that parallels more the phonological rules of spoken language than its syntax (Doupe & Kuhl, 1999; Berwick et al., 2010). Linear sequencing (e.g. word order) that does not rely on the computational system of grammar, can emerge in humans in simple improvised communication (Goldin-Meadow et al., 2008) through the direct interaction between the sensory-motor and conceptual systems (Langus & Nespor, 2010). It is therefore possible that the structural regularities in birdsong may also be generated in the absence of a modular computational system of grammar. However, it is important to note that these non-syntactic ordering regularities we observe in improvised gestural communication in humans do not extend to generative processes such as syntactic branching or recursion, and participants are incapable of exploiting them for describing structurally more complex situations (Langus & Nespor, 2010). In contrast, Bengalese finches can not only vary the number of repeated syllables but also their linear organization according to a finite-state grammar (Hosino & Okanoya, 2000). These processes by far surpass the structural regularities that humans can generate without applying syntactic rules. The structural flexibility in birdsong, even though generated according to a finite state grammar, may thus suggest that the computations necessary for sequencing the song elements are independent of the sensory-motor representations over which they operate.

(d) Acquisition

There are striking similarities also in the learning of birdsong and spoken language. The processes involved in acquiring birdsong and spoken language are instinctive and do not require explicit tutoring, a type of learning that has only been documented in humans (Gleitman & Wanner, 1982), songbirds (Marler & Slabbekoorn, 2004; Thorpe, 1958), various species of parrots (Marler & Slabbekoorn, 2004; Gramza, 1970; Todt, 1975), Anna hummingbirds (Baptista, 1990), sack-winged bats (Knörnschild, Behr, & von Helversen, 2006), a harbour seal (Ralls, Fiorelli, & Gish, 1985) and two elephants (Poole et al., 2005). Just like the first uttered words in infants emerge from non-verbal babbling also juvenile birds begin with a sub-song that is then molded into species-specific vocalization (Doupe & Kuhl, 1999). Both language and birdsong are learned best during a 'period of opportunity' early in development that is called 'sensitive period' (Hurford, 1991; Brainard, & Doupe, 2002). If con-specific input is not available during this window of opportunity, neither humans will fully master language (Fromkin et al., 1974), nor birds their song (Marler, 1970; Thorpe, 1958). However, if children acquire their mother tongue from inconsistent linguistic input (Senghas et al., 1997; Sandler et al., 2005) and juvenile birds their song from inconsistent song input (Fehér, 2009), both standardize their vocal communication according to species-specific characteristics, suggesting a strong genetic blue-print for both language and birdsong.

Young pre-verbal infants approach the speech stream equipped with a "toolkit" of simple signal-driven processes such as transitional probabilities between adjacent (Saffran et al., 1996) as well as non-adjacent syllables (Peña et al., 2002), algebraic rule generalizations (Marcus et al., 1999), repetition detection (Endress et al., 2005) and prosodic bootstrapping **(**Cutler et al., 1997; Langus et al., under revision**)**. For example, 7-month-old infants are able to extract from brief familiarization streams of artificial speech simple algebraic rules of the kind ABB (e.g. ga ti ti), to generalize these rules to novel syllables (e.g. wo fe fe), and to distinguish ABB rules from AAB (e.g. wo wo fe) and ABA rules (e.g. wo fe wo) (Marcus et al., 1999). While the precise nature of these signal driven computations is currently debated (see Endress, et al., 2005), they appear to be so simple that they are readily employed by sleeping neonates (Gervain et al., 2008), by rats (Murphy et al., 2008), by songbirds (Abe & Watanabe, 2011) and possibly even by honeybees (Giurfa et al., 2001). The fact that these mechanisms are shared with other non-human animals (c.f. Hauser et al., 2002; Endress et al., 2009) implies they are not language specific.

Song acquisition may parallel early language acquisition in several aspects (Doupe & Kuhl, 1999). For example, just like human infants, who rely on signal driven processes over chunks of speech (e.g. syllables) (Peña et al., 2002), also songbirds learn their song in chunks (e.g. segments of syllables from one or more adults' song) that can be rearranged during song acquisition to form a new song (Williams & Staples, 1992; Wheelwright et al., 2008). There is even some preliminary evidence, which suggests that song learning as well as singing requires birds to store and retrieve rule-related knowledge that may be acquired through signal-driven learning (Hultsch et al., 1999). However, to what extent birdsong acquisition parallels spoken language acquisition, and whether the processes mediating song and speech acquisition are similar, remains to be determined.

3. FoxP2[2] in language and birdsong

Language in humans emerged through qualitative and quantitative modifications of morphology that existed in our primate ancestors. Molecular mechanisms that shape such changes are e.g. heterotypy (altered gene products), heterochrony (altered timing of gene expression), heterotopy (altered spatial gene expression), and heterometry (altered amounts of expression). For example, eyes did evolve in many different ways in different species and did not evolve from a common ancestor eye, but the transcription factor Pax6 nevertheless plays a central role during eye formation across the entire phylogenetic tree (Fernald, 2006). More and more conserved molecular toolkits emerge, for example for learning and memory in flies, slugs and mice (Müller, 2006). Do conserved molecular networks play a role in learned vocal behavior such as spoken language and birdsong?

The link between the transcription factor FOXP2 and language was discovered because a mutated version of *FOXP2* was inherited in an autosomal dominant fashion by three generations of the so-called KE family (Lai et al., 2001). The language impairment is a form of Developmental Verbal Dyspraxia (DVD) and characteristically entails variable and inaccurate pronunciation of words, impaired repetition of long words, and also problems with processing complex syntax (Stackhouse, 1992; Lai et al., 2001; Watkins et al., 2002). *In vitro* and *in vivo* studies, including experiments with model organisms like mice, songbirds, fish and flies address the molecular and neural function of FoxP2 in different species (Fisher & Scharff, 2009; White et al., 2006). Results are consistent with a role both in neural development and neural function in circuits specialized for translating sensory information into motor behavior, a crucial aspect of spoken language. It is possible that the *FOXP2* gene plays a particularly important, specialized role in speech and language as a result of positive selection in the hominine lineage (Enard et al., 2002; Yu et al., 2009), but whether and how this is true is not known yet. Studying its molecular, cellular and behavioral function in animal models has already provided a glimpse at what type of molecular pathways may be particularly relevant to speech and language. This kind of research may in time offer biologically grounded and eventually testable hypotheses on how these human-specific traits may have evolved (Lai et al., 2001; Carroll, 2005; Scharff & Petri, 2011).

The FoxP2 protein belongs to a family of transcription factors that are characterized by a highly conserved Forkhead (Fox) domain that binds to distinct DNA sequences in the regulatory regions of its target genes. Binding to these target (or 'downstream') genes can either repress or activate their transcriptional activity (Shu et al., 2001; Vernes et al., 2007; Vernes et al., 2011). Additional FoxP2 protein domains provide further opportunities to interact with DNA of other genes and proteins, for instance dimerizing with two other FoxP family members, FoxP1 and FoxP4. In fact, dimerization of FoxP2 with other FoxPs is essential for transcriptional repression, at least in reporter-gene cell culture assays (Wang et al., 2003; Li et al., 2004).

Specific genes to which FOXP2 binds were identified in a human neuronal cell line (Vernes et al, 2007), in various human embryonic tissues (Spiteri et al., 2007) and in mice (Vernes et al., 2011). Many candidate genes from these studies are proposed to play an important role in neurodevelopment and neurotransmission, implying dysfunction of these pathways in patients with *FOXP2* mutations (Spiteri et al., 2007; Vernes et al., 2007, 2011). *CNTNAP2* is a downstream target of FOXP2 and of particular interest. Certain *CNTNAP2* sequence variants (single nucleotide polymorphisms, SNPs) are associated with language-delays in autistic children and are linked to core deficits of language impaired children as well as children with dyslexia (Vernes et al., 2008; Raskind et al., 2011). Interestingly, in songbirds *CNTNAP2* is differentially expressed in some of the song control nuclei (Panaitof et al., 2010), but whether FoxP2 regulates *CNTNAP2* has not been addressed in birds yet. These are the first data that show interesting similarities in songbirds and humans of molecular players interacting with FoxP2, pointing towards a potential deep homology between human speech and bird song.

In order to understand whether and how *FoxP2* plays a role in the evolution of vocal learning, it is necessary to understand how FoxP2 and its associated molecular network affects neuronal development and neural function in distinct species and during different stages of development. Konopka et al. (2009) found that expression levels of 116 genes differed quantitatively when they compared the chimp and the human *FoxP2* versions in a human neuronal cell line. Among this set of genes many were acting in pathways and were expressed in tissues relevant for speech and language.

In addition to understanding how FoxP2 regulates other genes' transcription ('downstream of FoxP2'), it is of equal importance to understand how the

---

[2] The nomenclature for Forkhead (Fox) genes follows Kaestner et al. (2000): human, FOXP2; mouse, Foxp2; and all other species, FoxP2. As per convention, genes and mRNA are italicized, proteins are not.

transcription of *FoxP2* is regulated ('upstream of FoxP2'). Evolutionary changes particularly in upstream regulatory regions are hypothesized to be particularly important in the course of evolution (Carroll, 2003). Two putative upstream transcriptional regulators of *FoxP2* are known, the transcription factor Lef1 in zebrafish (*Danio rerio*) (Bonkowsky et al., 2008) and endocannabionoids in zebra finch (*Taenigpygia guttata*) (Soderstrom & Luo, 2009). Whether cannabinoid signalling has direct or indirect effects on FoxP2 expression in the song system needs further investigation.

(a) *FoxP2* expression during brain development and in postnatal brains

There are three main ways in which the FOXP2 protein could affect language and speech in humans. It could be important for the formation of speech-circuits, it could be involved in the process of speech learning, and/or in the perception and/or the production of speech. *FOXP2* mRNA is already expressed in human fetal brains in the same regions that later develop morphological and functional abnormalities in affected patients with mutated *FOXP2* (Lai et al., 2003). *FOXP2* is therefore thought to play a role in establishing speech relevant circuits very early in development. Expression patterns in fetal brains of humans, other primates, different species of mice, different species of birds and fish of similar embryonic stages largely coincide, but interesting differences exist as well (Bonkowsky & Chien, 2005; Ferland et al., 2003; Haesler et al., 2004; Itakura et al., 2008; Schön et al., 2006; Shah et al., 2006; Shu et al., 2001; Takahashi et al., 2003, 2008; Teramitsu et al., 2004). Since FoxP2 is not exclusive to humans, it is very likely to act within similar pathways during the development of homologous brain regions in many vertebrates.

FoxP2 expression in songbirds varies during development as well as in different behavioral contexts in a basal ganglia structure called Area X. Young songbirds acquire their species-specific and individual-specific song via imitation of adult conspecifics. Area X needs to be intact for song learning to occur normally (Scharff & Nottebohm, 1991; Sohrabji, 1990). After song has been learned, Area X continues to be relevant for online monitoring of song (Scharff & Nottebohm, 1991; Sohrabji, 1990; Kao et al., 2005) and without Area X, normal song production deteriorates (Kobayashi et al., 2001). Juvenile male zebra finches consistently express 10-20% more *FoxP2* mRNA within Area X compared to the surrounding striatum during song learning (Haesler et al., 2004). Most other song control regions show very low levels of *FoxP2* expression (Haesler et al., 2004; Teramitsu et al., 2004). A further correlation between song plasticity and levels of *FoxP2* expression exists in canaries. *FoxP2* expression in Area X is low during the breeding season, when male canaries sing highly regular and stereotyped songs. Song becomes more variable and new syllables are incorporated after the breeding season, and concomitantly *FoxP2* in Area X is upregulated (Haesler et al., 2004). The different *FoxP2* mRNA levels in these experiments are not explainable by a direct relationship with recent singing activity.

Yet, singing also contributes to FoxP2 mRNA and protein levels, with some interesting differences. At around 75 days after hatching during the late phase of song development, singing undirected song for two hours is correlated with low *FoxP2* mRNA levels in Area X. This is also the case in adult finches, where *FoxP2* mRNA is lower after undirected singing than after female directed courtship singing (or no singing at all) (Teramitsu & White, 2006). In contrast, FoxP2 protein levels in Area X are low, both after either undirected or directed song (Miller et al., 2008). The nature of this difference is currently not clear. Interestingly, during the first two hours of daylight FoxP2 protein in Area X increases in non-singers, stressing that FoxP2 in songbirds is regulated by singing as well as other mechanisms (Haesler et al., 2004; Miller et al., 2008; Teramitsu et al., 2010).

The song circuitry in species of two other avian orders that learn their vocalizations by imitation, e.g. hummingbirds and parakeets, also expresses *FoxP2*. Vocal production learning may have evolved multiple times during evolution, once in the common ancestor of parrots and songbirds and once in the more distantly related hummingbirds (Hackett et al., 2008; Suh et al., 2011). An alternative hypothesis is however also possible. Vocal learning might have existed in a common ancestor but was lost subsequently in many extant lineages (Jarvis, 2004). When both scenarios are considered, one should bear in mind that for only a few species the absence of vocal production learning has been clearly demonstrated (Kroodsma & Konishi, 1991). There may very well be so far unrecognized intermediate phenotypes between so-called accurate imitative 'production' learning and 'usage' learning (Janik, 2000). More research into this issue is necessary, including the search for neural structures similar to those in the three vocal learning bird orders and their FoxP2 expression in these brain regions. These types of experiments are essential to describe the nature of universal deep molecular homologies relating to the neurobiology and the behavior of vocal production learning.

(b) Functional analysis of FoxP2

To establish whether FoxP2 affects complex learned vocalizations in songbirds, *FoxP2* was knocked-down in juvenile male zebra finches during the sensorimotor learning phase in the striatal nucleus Area X (Haesler et al., 2007). Reduced FoxP2 levels in Area X led to incomplete copying of the tutor's song; several of the tutor's syllables were not imitated at all, and others were copied inaccurately. Furthermore, knock-down birds sang their songs more variably from rendition to rendition (Haesler et al., 2007). Exaggerated variability of speech production is also a characteristic of people with *FOXP2* mutations (Watkins et al., 2002). Even though the current animal models do not allow to pinpoint the relative contribution of sensory, motor or sensorimotor integration to the specific impairments, a number of findings suggest that the deficits resulting from *FoxP2* knockdown are not restricted solely to motor performance, nor to sensory processing alone (Haesler et al., 2007; Teramitsu et al., 2010). In sum, experimentally reduced FoxP2 levels in striatal Area X of juvenile male zebra finches caused song impairments that phenotypically echo aspects of developmental verbal dyspraxia in humans: like patients with mutated *FOXP2*, birds with reduced FoxP2 levels fail to develop their full articulatory potential and produce a reduced set of vocal elements more variably than is species-typical.

Do these findings allow us to predict any function of FoxP2 at the neural level? Despite some important differences, the anterior forebrain pathway (AFP) of song learning birds echoes the mammalian cortico-basal ganglia-thalamo-cortical loops. Like the striatum in mammals, the striatal Area X in songbirds receives cortical glutamatergic afferents that synapse onto spiny neurons with histochemical and electrophysiological features very similar to those of mammalian medium spiny neurons. The cortical input to the spiny neurons of Area X of songbirds is also modulated presynaptically by midbrain dopaminergic input. However, Area X also contains aspiny, tonically active, fast firing GABAergic neurons similar to mammalian pallidal neurons (Farries et al., 2005). Electrophysiological recordings from Area X in singing birds reveal two types of these neurons, that differ in connectivity and firing patterns akin to the two different pallidal neuron types in primates (Goldberg et al., 2010). However, Area X within the songbird striatum has slightly different connectivity patterns than those of the surrounding striatum (Person et al., 2008). These differences could reflect the small, but interesting evolutionary modifications postulated for new traits, such as avian vocal learning.

(c) Evolution of *FoxP2*

The *FoxP2* gene is highly conserved, both in protein sequence and in brain expression pattern. The protein sequences of zebra finch and human FoxP2 differ by less than 2%, and the chimpanzee and human version are even more similar. Clearly the protein fulfils functions that are evolutionarily ancient and not limited to humans. Yet, there is also good evidence that the human *FOXP2* gene and some of its putative target genes are the result of strong positive selection (Enard et al., 2002; Zhang et al., 2002; Spiteri et al., 2007), not necessarily acting on the coding region of the gene (Ptak et al., 2009). The question is whether the difference in the human and chimp FOXP2 sequences have anything to do with the fact that humans speak and chimps do not. To address this, Enard et al (2010) genetically engineered a mouse that contains a humanized *FOXP2* version instead of its usual murine version. Pups of this mouse strain produce isolation calls that differ bioacoustically from those of control mice; other abnormalities include less exploratory behaviour, altered synaptic plasticity of striatal medium spiny neurons, lower dopamine levels in five brain regions including the frontal cortex and the caudate-putamen, and longer dendrites in cell culture. The humanized version of *FOXP2* introduced into the mouse genome was shown to specifically affect the cortico-basal ganglia circuits, but not the cortico-cerebellar circuits in a follow-up study (Reimers-Kipping et al., 2010). This highlights that a widely distributed protein like FoxP2 can have quite specific effects on neuro-development of particular circuits and thus function, contrary to the frequently cited misconception ‘FoxP2 cannot be important for something as specific as language if the protein also plays a role in the lung’.

The original estimate for the human specific *FOXP2* sequence was dated around 260,000 years ago, and therefore concomitant with the emergence of cultural artifacts that are thought to be indicative of concomitant language evolution. However, recent evidence has challenged this finding. Neanderthals may have already possessed the human-like *FOXP2* version, even though their lineage diverged from the one leading to modern man already approximately 300,000-400,000 years ago (Krause et al., 2007). It cannot be excluded that genes were exchanged between Neanderthals and humans or that Neanderthal samples were contaminated with modern human DNA (Coop et al., 2008). Furthermore, genomic evidence suggests that the human *FOXP2* mutations may have occurred 1.8 or 1.9 million years ago, around the time when the *genus homo* emerged (Diller & Cann, 2009). This may mean that the human *FOXP2* mutations occurred either before the emergence of spoken language or that a more primitive form of spoken language may have preceded modern human language capacities.

Thus, to understand the dynamics of *FoxP2*’s expression in different species is one of the great challenges to understand the gene’s function and its evolutionary role in vocal learning. There are *a* number of open questions that need to be addressed in the future: (1) How are FoxP2 expression levels affected by non-coding DNA sequence changes in different species? (2) How do coding changes affect the structure and the interaction of the protein with the DNA of its targets and other proteins? (3) How does the differential expression of cofactors and other proteins that interact with FoxP2 influence its function? (4) How do FoxP2 expression levels respond to internal and external influences? All of these factors might be important for evolutionary change. Songbirds for example are a fruitful model system to explore heterochrony. Age related differences, seasonal changes and differences dependent on different singing styles exist in striatal Area X (Haesler et al., 2004; Teramitsu et al., 2004, 2006, 2010). In the future, it will be interesting to follow along these lines different species of songbirds that vary in the timing of song learning that sing during those behavioral contexts for which FoxP2 expression has not been tested yet, and that show differences especially in adult song plasticity.

4. Conclusions

DNA sequence changes in the regulatory regions of certain genes can alter the amount as well as the timing, or the place of the gene’s expression in the course of evolution (Carroll, 2008). Likewise, coding changes can alter gene products, leading to different functions (Hoekstra & Coyne, 2007). Thus, both factors can result in changes to neural circuitry, as amply attested by differences in neuroanatomy among different species. Whether *FoxP2* played such a role in bringing about specifically those circuit changes that facilitated the emergence of human language is not clear yet. But since a complex behavior like language (as well as birdsong) is bound to be a polygenic trait, other, so far unknown, genes need to act together with FOXP2.

Alternatively, other genes brought about the circuit changes required for vocal learning. Subsequently, FoxP2, which already functioned in the precursor circuits, either acquired new importance because it operated in a new environment, or the gene changed its function. In songbirds, song nuclei are embedded in regions that are active during stereotyped motor behaviors, like hopping and walking (Feenders et al., 2008). This suggests that if during evolution Area X developed out of existing FoxP2 expressing striatal territories in ancestral birds that were non-vocal learners, *FoxP2* expression in Area X may have become useful for sensory motor integration necessary for the precise timing of vocal gestures. In addition, *FOXP2* may have mutated in humans to become human specific, and this might have further affected neural development and transmission in cortico-striatal circuits relevant for speech and language. This would be a two-hit scenario for the role of *FOXP2* in the evolution of language, circuit changes predating gene function changes. From the post-natal studies in birds it is clear that FoxP2 plays a role in neural plasticity of certain circuits. However, whether this is true for brain circuits that are relevant for vocal learning in humans as well as in songbirds is not clear.

What is clear is that both spoken language and birdsong are complex communicative behaviors that require auditory-guided vocal motor learning. Because the comparative studies have primarily focused on birdsong as a phonological phenomenon, the behavioral as well as genetic and molecular evidence for parallels between birdsong and spoken language is strongest for sensorimotor processes that underlie song production, perception and acquisition. The evidence suggests that underlying neural and genetic substrates as well as environmental input play a role in the acquisition of species-specific vocalizations. The vast majority of the data pertaining to the expression of the *FoxP2* gene in humans and songbirds as well as its functional role in spoken language and birdsong, suggest that the FoxP2 transcription factor and its target genes may function in a similar manner to conserved molecular toolkits that have been found for shaping morphology of form. Even though the precise nature of the differences between the FoxP2 in humans and songbirds is not completely understood, comparative studies that sketch out the similarities and differences between spoken language and birdsong are necessary to complement the molecular and genetic studies that aim at determining the function of FoxP2 in birdsong and spoken language.

Finally, recent attempts to compare the syntax of birdsong and spoken language suggest that the parallels between the communicative behaviors of humans and songbirds are not limited to sensorimotor processes (Bolhuis et al., 2010; Berwick et al., 2011). However, our knowledge of the possible conceptual-intentional processes underlying bird vocalizations is almost non-existent. Thus many questions remain unanswered: To what extent do the conceptual-intentional processes and syntactic computations exist in birdsong? Can birdsong express different affective ('emotional') states? Can modularity of birdsong be usefully compared to modular components of spoken language, e.g. syntax and semantics? To what extent are these behaviors genetically encoded by conserved molecular toolkits as sensorimotor processes appear to be? Further research will have to answer these questions through both behavioral comparisons as well as investigations of their molecular and genetic basis.

References:

Abe, K., Watanabe, D. (2011) Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat Neurosci*. 14, 1067-74

Anderson, S. E., Dave, A., & Margoliash, D. (1996). 'Template-based automatic recognition of birdsong syllable from continuous recordings', *Journal of Acoustic Society of America*, 100, 1209-1219.

Aronov, D., Andalman, A. S., & Fee, M.S. (2008). 'A specialized forebrain circuit for vocal babbling in the juvenile songbird', *Science*, 320, 630-634.

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

Bauer, L. (1983). *English Word-formation*. Cambrdige: Cambrdige University Press.

Beckman, M. & Pierrehumbert, J. (1986). 'Intonational Structure in Japanese and English', *Phonology Yearbook*, 3, 15-70.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L., Mehler, J. (1988). 'An Investigation of Young Infants' Perceptual Representations of Speech Sounds', *Journal of Experimental Psychology : General*, 117, 21-33.

Bertoncini, J., Floccia, C., Nazzi, T., Mehler, J. (1995). 'Morae and Syllables: Rythmical Basis of Speech Representations in Neonates', *Language and Speech,* 38, 311-329.

Bertossa, R., (2011) 'Morphology and behaviour: functional links in development and evolution'. *Philosophical Transactions of the Royal Society B*, 366, 2056–2068.

Berwick, R. C., Okanoya, K., Beckers, G. J. L., Bolhuis, J. J. (2011). 'Songs to syntax: the linguistics of birdsong', *Trends in Cognitive Science*, 15, 113-121.

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

Bickerton, D. (2010). 'On two incompatible theories of language evolution', in R. Larson, V. Déprez, & H. Yamakido, (eds.), *The Evolution of Human Language: Biolinguistic Perspectives*. Cambridge: Cambridge University Press.

Bijeljac-Babic, R., Bertoncini, J., Mehler, J. (1993). 'How Do Four-Day-Old Infants Categorize Multisyllabic Utterances?' *Developmental Psychology,* 29, 711-723.

Bion, R. A. H., Benavides, S., & Nespor, M. (2011). Acoustic markers of prominence influence adults' and infants' memory of speech sequences. *Language & Speech,* 54(1), 123-140.

Bohn, K. M., Schmidt-French, B., Schwartz, C., Smotherman, M., & Pollak, G. D. (2009). 'Versatility and Stereotypy of Free-Tailed Bat Songs', *Plos One*, 25, 4.

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

Bolinger, D. (1989). Intonation and its Uses: Melody in Grammar and Discourse, Stanford University Press, Stanford.

Bonkowsky, J. L. & Chien, C. B. (2005). 'Molecular cloning and developmental expression of foxP2 in zebrafish', *Developmental Dynamics*, 234, 740-746.

Bonkowsky, J. L., Wang, X., Fujimoto, E., Lee, J. E., Chien, C. B., & Dorsky, R. I. (2008). 'Domain-specific regulation of foxP2 CNS expression by lef1', *BMC Developmental Biology*, 8, 103.

Brainard, M. S. & Doupe, A. J. (2002). 'What songbirds teach us about learning', *Nature*, 417, 351–358.

Caceres, M., Lachuer, J., Zapala, M.A., Redmond, J.C., Kudo, L., Geschwind, D.H,, Lockhart, D. J., Preuss, T. M., & Barlow, C. (2003). 'Elevated gene expression levels distinguish human from non-human primate brains', *Proceedings of the National Academy of Sciences of the United States of America*, 100, 13030-13035.

Carroll, S. B. (2003). ‘Genetics and the making of *Homo sapiens*’, *Nature*, 422, 849-857.

Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS Biol.* 3, e245

Carroll, S. B. (2008). ‘Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution’, *Cell*, 134, 25-36.

Catchpole, C. K. & Slater, P. J. B. (2008). *Bird Song: Biological Themes and Variations*. Cambridge: Cambridge University Press.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1980). ‘Principles and parameters in syntactic theory’, in N. L. Hornstein & D. Lightfoot (eds.), *Explanation in linguistics: The logical problem of language acquisition*. London and New York: Longman.

Chomsky, N. (1995). *Minimalist Program*. Cambridge: MIT Press.

Coop, G., Bullaughey, K., Luca, F., & Przeworski, M. (2008). ‘The timing of selection at the human FOXP2 gene’, *Molecular Biology and Evolution*, 25, 1257-1259.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). ‘Prosody in the comprehension of spoken language: a literature review’, *Language and Speech*, 40, 141–201.

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

Dehaene-Lambertz, G.Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). ‘Functional organization of perisylvian activation during presentation of sentences in preverbal infants’, *Proceedings of the National Academy of Sciences of the United States of America,* 103, 14240–14245.

Diller, K. C. & Cann, R. L. (2009). Evidence Against a Genetic-Based Revolution in Language 50,000 Years Ago. In Botha, R. & Knight, C. (eds.). *The Cradle of Language*,

Dooling, R. & Searcy, M. (1980). ‘Early perceptual selectivity in the swamp sparrow’, *Developmental Psychobiology*, 13, 499–506.

Doupe, A.J. & Konishi, M. (1991). ‘Song-selective auditory circuits in the vocal control system of the zebra finch’, *Proceedings of the National Academy of Sciences of the United States of America,* 88, 11339–11343.

Doupe, A. J. & Kuhl, P. K. (1999). ‘Birdsong and human speech: common themes and mechanisms’, *Annual Review of Neuroscience*, 22, 567-631.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., Vigorito, J. (1971). ‘Speech perception in infants’, *Science*, 171, 303-306.

Elemans, C., Laje, R., Mindlin, G.B. & Goller, F. (2010). ‘Smooth Operator: avoidance of subharmonic mechanical mechanisms simplifies song motor control in adult zebra finches.’ *J Neurosci*, 30:13246-13250.

Emery, N. J. & Clayton, N. S. (2004). ‘The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes’. *Science*, 306, 1903-1907.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., Monaco, A. P., & Pääbo, S. (2002). ‘Molecular evolution of FOXP2, a gene involved in speech and language’, *Nature*, 418, 869-872.

Enard, W., Gehre, S., Hammerschmidt, K., Holter, S. M., Blass, T., Somel, M., et al. (2010). ‚A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice.’ *Cell*, 137, 961-971.

Endress, A.D., Nespor, M. & Mehler, J.(2009). ‘Perceptual and Memory Constraints on Language Acquisition’, *Trends in Cognitive Science*, 13, 348-353.

Endress, A. D., Scholl, B. J., Mehler, J. (2005). ‘The role of salience in the extraction of algebraic rules’, *Journal of Experimental Psychology: General*, 134, 406-419.

Fee, M. S., Shraiman, B., Pesaran, B., & Mitra, P. P. (1998). ‘The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird’, *Nature*, 395, 67–71.

Farries, M. A. Ding, L. & Perkel, D. J. (2005) ‚Evidence for ‘‘direct’’ and ‘‘indirect’’ pathways through the songsystem basal ganglia.’ *J. Comp. Neurol.* 484, 93–104.

Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., Kazuhiro, W., Mouritsen, H., & Jarvis, E. D. (2008). ‘Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin’, *PLoS One*, 3:e1768.

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*, 459, 564-568.

Ferland, R. J., Cherry, T. J., Preware, P. O., Morrisey, E. E., & Walsh, C. A. (2003). “Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain’, *Journal of Comparative Neurology*, 460, 266-79.

Fernald, R. D. (2006). ‘Casting a genetic light on the evolution of eyes’, *Science*, 313, 1914-1918.

Fisher SE, Scharff C. (2009). ‘FOXP2 as a molecular window into speech and language.’ *Trends Genet.* 25, 166-77.

Fitch, W.T. (2009). ‘The Biology and Evolution of Language: “Deep Homology” and the Evolution of Innovation’, in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*. Cambridge: MIT Press.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge: MIT Press.

Frisch, K. (1967). *The Dance Language and Orientation of Bees*. Cambridge: Harvard University Press.

Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., & Rigler, M. (1974). The development of language in Genie: A case of language acquisition beyond the critical period. *Brain and Language*, 1, 81-107.
Gaunt, A. S. (1987). 'Phonation', in T. J. Seller (ed.). *Bird Respiration*, Boca Raton, FL: CRC.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., Mehler, J. (2008). 'The neonate brain detects speech structure', *Proceedings of the National Academy of Sciences of the United States of America*, 105, 14222-14227.

Gilbert, S. F., Opitz, J.M., & Raff, R.A. (1996). 'Resynthesizing evolutionary and developmental biology', *Develomental Biology*, 173, 357-72.

Giurfa, M., Shaowu, Z., Arnim, J., Randolf, M., & Mandyam, S. V. (2001). 'The concepts of 'sameness' and 'difference' in an insect', *Nature*, 410, 930-933.

Goldberg, J. H., Adler, A., Bergman, H., & Fee, M. S. (2010). 'Singing-related neural activity distinguishes two putative pallidal cell types in the songbird basal ganglia: comparison to the primate internal and external pallidal segments', *Journal of Neuroscience*, 30, 7088-1098.

Goldin-Meadow, S., So, W. C., Ozyürek, A., & Mylander, C. (2008). 'The natural order of events: How speakers of different languages represent events nonverbally', *Proceedings of the National Academy of Sciences of the United States of America*, 105, 9163–9168.

Goller, F. & Larsen, O. N. (1997). 'A new mechanism of sound generation in songbirds', *Proceedings of the National Academy of Sciences of the United States of America,* 94, 14787–14791.

Gramza, A. F. (1970). 'Vocal mimicry in captive budgerigars (*Melopsittacus undulatus*)', *Zeitschrift für Tierpsychologie*, 27, 971-983.

Griesser, M. (2008). *Referential calls signal predator behavior in a group-living bird species*. Current Biology, 18, 69-73.

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D., & Scharff, C. (2004). 'FoxP2 expression in avian vocal learners and non-learners', *Journal of Neuroscience*, 24, 3164-3175.

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). 'Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X', *PLoS Biol*ogy, 5, e321.

Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., Braun, M. J., Chojnowski, J. L., Cox, W. A., Han, K. L., Harshman, J., Huddleston, C. J., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Steadman, D. W., Witt, C. C., & Yuri, T. (2008). 'A phylogenomic study of birds reveals their evolutionary history', *Science*, 320, 1763-1768.

Hahnloser, R. H. R. & Kotowicz, A. (2010). 'Auditory representations and memory in birdsong learning' *Current Opinion in Neurobiology*, 20, 332-339.

Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002). 'The faculty of language: what is it, who has it, and how did it evolve?' *Science*, 298, 1569-79.

Hauser, M. D. & Konishi, M. (1999). *The Design of Animal Communication*. Cambridge: MIT Press.

Hay, J., & Diehl, R. (2007). 'Perception of rhythmic grouping: Testing the iambic/trochaic law', *Perception & Psychophysics*, 69, 113-122.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: The University of Chicago Press.

Hayes, B. (1989). 'The prosodic hierarchy in meter', in P. Kiparsky & G. Youmans (Eds.), *Phonetics and phonology, Vol 1: Rhythm and meter*, San Diego: Academic Press,

Hickok, G. & Poeppel, D. (2000). 'Towards a functional neuroanatomy of speech perception', *Trends in Cognitive Sciences*, 4, 131–138.

Hockett, C. F. (1960). 'The origin of speech', *Scientific American*, 203, 89-96.

Hoekstra, H. E. & Coyne, J. A. (2007). 'The locus of evolution: evo devo and the genetics of adaptation', *Evolution*, 61, 995-1016.

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

Hultsch, H., Mundry, R. & D. Todt (1999). Learning, representation & retrieval of rule-related knowledge in song systems. In: Friederici, A. & Menzel, R. (eds). Learning: Rule Extraction and Representation. Berlin & New York: Walter de Gruyter, pp 87-99.

Hultsch, H. & Todt, D. (1989). 'Memorization and Reproduction of Songs in Nightingales (*Luscinia megarhynchos*) - Evidence for Package Formation', *Journal of Comparative Physiology*, 165, 197-203.

Hultsch, H. & Todt, D. (1999). 'Learning, representation & retrieval of rule-related knowledge in song systems', in A. Friederici & R. Menzel (eds.) *Learning: Rule Extraction and Representation*. Berlin: Walter de Gruyter.

Hultsch, H. & Todt, D. (2001). 'Developmental trajectories of complex signal systems in animals: the model of bird song', in: J. W. B. Höhle (ed.), *Approaches to Bootstrapping*. Amsterdam: John Benjamin.

Hurford, J.R. (2008). 'The evolution of human communication and language in Sociobiology of communication: an interdisciplinary perspective', in P. d'Ettorre & D. P. Hughes (eds.), *Sociobiology of Communication an interdisciplinary perspective*. Oxford: Oxford University Press.

Hurford, J. R. (1991). 'The Evolution of the critical period for language acquisition', *Cognition*, 40, 159–201.

Immelmann K. (1969). 'Song development in the zebra finch and other estrildid finches', in R. A. Hinde (ed.) *Bird Vocalizations*, London: Cambridge University Press.

Imeda, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). 'Infant speech perception activates Broca's area: a developmental magnetoencephalography study', *Neuroreport*, 17, 957-962.

Indefrey, P. & Cutler, A. (2004). 'Prelexical and Lexical Processing in Listening', in M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, Cambridge: MIT Press.

Itakura, T., Chandra, A., Yang, Z., Xue, X., Wang, B., Kimura, W., Hikosaka, K., Inohaya, K., Kudo, A., Uezato, T., & Miura, N. (2008). 'The medaka FoxP2, a homologue of human language gene FOXP2, has a diverged structure and function', Journal of Biochemistry, 143, 407-416.

Janik, V. M. & Slater, P. J. B. (2000). 'The different roles of social learning in vocal communication', *Animal Behavior*, 60, 1-11.

Jarvis, E. D. (2004). 'Learned birdsong and the neurobiology of human language', *Annals of the New York Academy of Sciences*, 1016, 749-177.

Kaestner, K.H., Knochel, W., & Martinez, D.E. (2000). Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev*. 14, 142–146.

Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). 'Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song', *Nature*, 433, 638-643.

Kipper, S., Mundry, R., Hultsch, H., & Todt, D. (2004). 'Long-term persistence of song performance rules in nightingales (*Luscinia megarhynchos*): A longitudinal field study on repertoire size and composition', *Behaviour*, 141, 371-390.

Klima, E. & Bellugi, U. (1979). *The Signs of Language*. Harvard University Press.

Kobayashi, K., Uno, H., & Okanoya, K. (2001). 'Partial lesions in the anterior forebrain pathway affect song production in adult Bengalese finches', *Neuroreport*, 12, 353-358.

Kojima, S. & Doupe, A. (2011). 'Social performance reveals unexpected vocal competency in young songbirds', *Proceedings of the National Academy of Sciences of the United States of America.* doi: 10.1073/pnas.1010502108

Konopka, G., Bomar, J. M., Winden, K., Coppola, G., Jonsson, Z.O., Gao, F., Peng, S., Preuss, T. M., Wohlschlegel, J. A., & Geschwind, D. H. (2009). 'Humanspecific transcriptional regulation of CNS development genes by FOXP2', *Nature*, 462, 213-217.

Knudsen, D. P. & Gentner, T. Q. (2010). 'Mechanisms of song perception in oscine birds', *Brain and Language*, 115, 59–68.

Knörnschild, M., Behr, O., & von Helversen, O. (2006). 'Babbling behavior in the sac-winged bat (*Saccopteryx bilineata*)', *Naturwissenschaften*, 93, 451-454.

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., Hublin, J. J., Hänni, C., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A., Pääbo, S., (2007). 'The derived FOXP2 variant of modern humans was shared with Neandertals', *Current Biology*, 17, 1908-1912.

Kreutzer, M., Beme, I., Vallet, E., Kiosseva, L. (1999). 'Social stimulation modulates the use of the 'A' phrase in male canary songs', *Behaviour*, 136, 1325–1334.

Kroodsma, D. E. & Byers, B. E. (1991). 'The function(s) of bird song', *American Zoology*, 31, 318-328.

Kroodsma, D. E. & Konishi, M. (1991). 'A Suboscine Bird (Eastern Phoebe, *Sayornis phoebe*) Develops Normal Song without Auditory-Feedback', *Animal Behavior*, 42, 477-487.

Kroodsma, D. E. & Parker, L. D. (1977). 'Vocal Virtuosity in Brown Thrasher', *Auk*, 94, 783-785.

Kumar, S. & Hedges, S. B. (1998). 'A molecular timescale for vertebrate evolution', *Nature*, 392, 917–920.

Ladd, D. R. (1996). *Intonational Phonology*, Cambridge University Press.

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

Lai, C. S., Gerrelli, D., Monaco, A. P., Fisher, S. E., & Copp, A. J. (2003). 'FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder', *Brain*, 126, 2455-2462.

Langmore, N. E. (1998). 'Functions of duet and solo songs of female birds', *Trends in Ecology and Evolution*, 12, 136-140.

Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (in press). ‘Can prosody be used to discover hierarchical structure in continuous speech’, *Journal of Memory and Language*.

Langus, A. & Nespor, M. (2010). ‘Cognitive systems struggling for word order’, *Cognitive Psychology*, 60, 291–318.

Lehiste, I. (1966). *Consonant Quantity and Phonological Units in Estonian*, Indiana University Publishers.

Lehiste, I. (1970). *Suprasegementals*. Cambridge: MIT Press.

Lenneberg, E. (1967). *Biological Foundations of Language*. New York: John Wiley & Sons.

Lewicki, M. S. & Konishi, M. (1995). ‘Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order’, *Proceedings of the National Academy of Science USA,* 92, 5582-5586.

Li, S., Weidenfeld, J., & Morrisey, E.E. (2004). ‘Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions’, *Molecular Cell Biology*, 24, 809-822.

Lieberman, P. (1967). *Intonation, Perception, and Language*, MIT Press, Cambridge.

Lombardino, A. J., Hertel, M., Li, X. C., Haripal, B., Martin-Harris, L., Pariser, E., Nottebohm, F. (2006). ‘Expression profiling of intermingled long-range projection neurons harvested by laser capture microdissection’, *Journal of Neuroscience Methods*, 157, 195-207.

Marcus, G. F., Vijayan, S., Bandi Rao, S., and Vishton, P. M. (1999). ‘Rule-learning in seven-month-old infants’, *Science*, 283, 77-80.

Margoliash, D. (1983). ‘Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow’, *Journal of Neuroscience,* 3, 1039–1057.

Margoliash, D. (1986). ‘Preference for autogenous song by auditory neurons in a song sys- tem nucleus of the white-crowned sparrow’, *Journal of Neuroscience,* 6, 1643–1661.

Margoliash, D. & Konishi, M. (1985). ‘Auditory representation of autogenous song in the song system of white-crowned sparrows’, *Proceedings of the National Academy of Science USA,* 82, 5997–6000.

Marler, P. (1970). ‘A comparative approach to vocal learning: song development in whitecrowned sparrows’, *Journal of Comparative and Physiological Psychology*, 71, 1–25.

Marler, P. & Slabbekoorn, H. W. (2004). *Nature's music: the science of birdsong*. Amsterdam: Elsevier Academic.

Mehler, J. & Bertoncini, J. (1981). ‘Syllables as Units in Infant Perception’, *Infant Behavior and Development*, 4, 271-284.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., Amiel-Tison, C. (1988). ‘A Precursor of Language Acquisition in Young Infants’, *Cognition*, 29, 143-178.

Mello, C. V., Vicario, D. S., & Clayton, D. F. (1992). ‘Song presentation induces gene expression in the songbird forebrain’, *Proceedings of the National Academy of Science USA*, 89, 6818-6822.

Miller, J. E., Spiteri, E., Condro, M. C., Dosumu-Johnson, R. T., Geschwind, D. H., & White, S. A. (2008). ‘Birdsong decreases protein levels of FoxP2, a molecule required for human speech’, *Journal of Neurophysiology*, 100, 2015-2025.

Mozziconacci, S. (2002). ‘Prosody and emotions’, *Proceedings of Speech Prosody 2002*, Aix-en-Provence.

Murphy, R. A., Mondragon, E., & Murphy, V. A. (2008). ‘Rule learning by rats’, *Science*, 319, 1849-1851.

Müller, U. (2006). ‘Memory: Cellular and molecular networks’, *Cellular and Molecular Life Sciences*, 63, 961-962.

Nazzi, T., Bertoncini, J. Mehler, J. (1998). ‘Language Discriminationby Newborns. Towards un understanding of the role of rhythm’, *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1-11.

Nelson, D.A. & Marler, P. (1993). ‘Innate recognition of song in white-crowned sparrows: a role in selective vocal learning?’ *Animal Behavior*, 46, 806–808.

Nespor, M., & Vogel, I. (2008). *Prosodic Phonology*. Berlin: Mouton de Gruyter. 1st edition 1986. Dordrecht: Foris.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio,* 2, 1-29.

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

Panaitof, S. C., Abrahams, B. S., Dong, H., Geschwind, D. H., & White, S.A. (2010). ‘Language-related Cntnap2 gene is differentially expressed in sexually dimorphic song nuclei essential for vocal learning in songbirds’, *Journal of Comparative Neurology*, 518, 1995-2018.

Peña, M., Bonatti, L., Nespor, M., Mehler J. (2002). ‘Signal-Driven Computations in Speech Processing’, *Science*, 298, 604-607.

Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Kiozumi, H., Bouquet, F., & Mehler, J. (2003). ‘Sounds and silence: an optical tomography study of language recognition at birth’, *Proceedings of the National Academy of Science USA*, 100, 11702–11705.

Peperkamp, S. & Mehler, J. (1999). ‘Signed and Spoken Language: A Unique Underlying System?’ *Language and Speech*, 42, 333-346.

Perani, D., Saccuman, M.C., Scifo, P., Awander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G. & Friederici, A.D. (2011). ‘Neural language networks at birth’, *Proceedings of the National Academy of Science USA*, Sep 6 Epub.

Person, A. L., Gale, S. D., Farries, M. A., & Perkel, D. J. (2008). ‘Organization of the songbird basal ganglia, including area X’, *Journal of Comparative Neurology*, 508, 840-866.

Pierrehumbert, J., & Hirschberg, J. (1990). ‘The meaning of Intonational contours in the interpretation of discourse’, in P. Cohen, J. Morgan, & M. Pollack (eds.), *Intentions in communication*. Cambridge, MA: The MIT Press.

Pinker, S. (1991). ‘Rules of Language’, *Science*, 253, 530-535.

Pinker, S. & Jackendoff, R. (2005). The faculty of language: what’s special about it? *Cognition* 95, 201–236.

Poeppel, D. (2001). ‘Pure word deafness and the bilateral processing of the speech code’, *Cognitive Science*, 21, 679–693.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). ‘Elephants are capable of vocal learning’, *Nature*, 434, 455-456.

Ptak, S. E, Enard, W., Wiebe, V., Hellmann, I., Krause, J., Lachmann, M., & Pääbo, S. (2009). ‘Linkage Disequilibrium Extends Across Putative Selected Sites in FOXP2’, *Molecular Biology and Evolution*, 26, 2181-1284.

Ralls, K., Fiorelli, P., & Gish, S. (1985). ‘Vocalizations and Vocal Mimicry in Captive Harbor Seals (*Phoca vitulina)*’, *Canadian Journal of Zoology*, 63, 1050-1056.

Raskind, B. P., Matsushita, W. H., Lisowski, M., Vu, M., Berninger, T., Wijsman, V. W., Brkanac, E. M. (2011). “Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample”, *Journal of Neurodevelopmental Disorders*, 3, 39-49.

Reimers-Kipping, S., Hevers, W., Pääbo, S., & Enard, W. (2010). ‘Humanized Foxp2 specifically affects cortico-basal ganglia circuits’, *Neuroscience*, 175, 75-84.

Rendall, D., Owren, M. J., & Ryan, M. J. (2009). ‘What do animal signals mean?’ *Animal Behaviour*, 78, 233-240.

Robinson, G. E., Fernald, R. D., & Clayton, D. F. (2008). ‘Genes and social behavior’, *Science*, 322, 896-900.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). ‘Statistical learning by 8-month-old infants’, *Science*, 274, 1925-1928.

San, D. (2007). *The Phonology of Standard Chinese*. Oxford University Press.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). ’The emergence of grammar: Systematic structure in a new language’, *Proceedings of the National Academy of Sciences of the United States of America*, 102, 2661–2665.

Sasahara, K. & Ikegami, T. (2007). ‘Evolution of birdsong syntax by interjection communication’, *Artificial Life*, 13, 259–277.

Scharff, C. & Nottebohm, F. (1991). ‘A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning’, *Journal of Neuroscience*, 11, 2896-2913.

Scharff, C. & Petri, J. (2011). Evo devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society B*, 366, 2124-2140.

Schön, C., Wochnik, A., Rossner, A., Donow, C., & Knochel, W. (2006). ‘The FoxP subclass in *Xenopus laevis* development’, *Development Genes and Evolution*, 216, 641-646.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge: MIT Press.

Selkirk, E. (1995). “Sentence prosody: Intonation, stress and phrasing”, in J. Goldsmith, ed., ‘The Handbook of Phonological Theory’, Blackwell, Cambridge, MA & Oxford, pp. 550–569.

Seyfarth, R.M., Cheney, D.L. & Marler, P. (1980). ‚Monkey responses to three different alarm calls: Evidence for predator classification and semantic communication’. *Science,* 210, 801-803.

Senghas, A., Coppola, M., Newport, E. L., & Supalla, T. (1997). Argument structure in Nicaraguan sign language: The emergence of grammatical devices. In E. Hughes & A. Greenhill (eds.), *Proceedings of the Boston university conference on language development*, Somerville: Cascadilla Press.

Shah, R., Medina-Martinez, O., Chu, L. F., Samaco, R. C., & Jamrich, M. (2006). ‘Expression of FoxP2 during zebrafish development and in the adult brain’, *International Journal of Developmental Biology*, 50, 435-8.

Shao, J. & Diamond, M. I. (2007). ‘Polyglutamine diseases: emerging concepts in pathogenesis and therapy’, *Human Molecular Genetics*, 16, R115-23.

Shu, W., Yang, H., Zhang, L., Lu, M. M., & Morrisey, E.E. (2001). ‘Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors’, *Journal of Biological Chemistry*, 276, 27488-27497.

Shubin N, Tabin C, Carroll S (August 1997). "Fossils, genes and the evolution of animal limbs". *Nature* **388** (6643): 639–48.

Slater, P. J. B., Eales, L. A., & Clayton, N. S. (1988). ‘Song learning in zebra finches: progress and prospects’, in J. S. Rosenblatt, C. Beer, M. C. Busnel, & P. J. B. Slater (eds.). *Advances in the Study of Behavior*. Harcourt.

Slobodchikoff, C. N., Paseka, A., & Verdolin, J.L. (2009). ‘Prairie dog alarm calls encode labels about predator colors’, *Animal Cognition*, 12, 435-9.

Soderstrom, K. & Luo, B. (2009). ‘Late-postnatal cannabinoid exposure persistently increases FoxP2 expression within zebra finch striatum’, Developmental Neurobiology, 70, 195-203.

Sohrabji, F., Nordeen, E. J., & Nordeen, K. W. (1990). ‘Selective Impairment of Song Learning Following Lesions of a Forebrain Nucleus in the Juvenile Zebra Finch’, *Behavioral and Neural Biology*, 53, 51-63.

Solis, M. M., Brainard, M. S., Hessler, N. A. & Doupe, A. J. (2000). ‘Song selectivity and sensorimotor signals in vocal learning and production’, *Proceedings of the National Academy of Sciences of the United States of America,* 97, 11836–11842.

Sossinka, R. & Böhner, J. (1980). ‘Song Types in the Zebra Finch *Poephila guttata castanotis*’, *Zeitschrift für Tierpsychologie*, 53, 123-32.

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou J, Vernes, S. C., Fisher, S. E., Ren, B., & Geschwind, D. H. (2007). ‘Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain’, *American Journal of Human Genetics*, 81, 1144-1157.

Stackhouse, J. (1992) ‚Developmental verbal dyspraxia. I: a review and critique.’ Eur. *J. Disord. Commun.* 27, 19–34.

Stern, D.L. & Orgogozo, V. (2009). ‘Is genetic evolution predictable?’ *Science*, 323, 746-751.

Stevens, K. N. (1994). ‘Scientific substrates of speech production’, in F. D. Minifie (ed.). *Introduction to Communication Sciences and Disorders*, San Diego: Singular.

Stulp, G., Emery, N.J., Verhulst, S., Clayton, N.S. 2009 ‘Western scrub-jays conceal auditory information when competitors can hear but cannot see’. *Biol Lett*. 5, 583-5.

Suh, A., Paus, M., Kiefmann, M., Churakov, G., Franke, F.A., Brosius, J., Kriegs, J.O., Schmitz, J. (2011). ’Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds.’ *Nat Commun*. 2, 443

Suthers, R. A. (1997). ‘Peripheral control and lateralization of birdsong’. *Journal of Neurobiology*, 33, 632– 52.

Suzuki, R., Buck, J. R., & Tyack, P.L. (2006). ‘Information entropy of humpback whale songs’, *Journal of the Acoustic Society of America*, 119, 1849-1866.

Takahashi, K., Liu, F. C., Hirokawa, K., & Takahashi, H. (2003). ‘Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum’, *Journal of Neuroscience Research*, 73, 61-72.

Takahashi, K., Liu, F. C., Oishi, T., Mori, T., Higo, N., Hayashi, M., Hirokawa, K., & Takahashi, H. (2008). ‘Expression of FOXP2 in the developing monkey forebrain: comparison with the expression of the genes FOXP1, PBX3, and MEIS2’, *Journal of Comparative Neurology*, 509, 180-9.

Teramitsu, I., Kudo, L. C., London, S. E., Geschwind, D. H., & White, S. A. (2004). ‘Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction’, *Journal of Neuroscience*, 24, 3152-3163.

Teramitsu, I. & White, S. A. (2006). ‘FoxP2 regulation during undirected singing in adult songbirds’, *Journal of Neuroscience*, 26, 7390-7394.

Teramitsu, I., Poopatanapong, A., Torrisi, S., & White, S. A. (2010). ‘Striatal FoxP2 is actively regulated during songbird sensorimotor learning’, *PLoS One*, 5:e8548.

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.

Thorphe, W. H. (1974). *Animal Nature and Human Nature*. London: Methuen.

Todt, D. (1975). ‘Social learning of vocal patterns and modes of their applications in Grey Parrots’, *Zeitschrift für Tierpsychologie*, 39, 178-188.

Trout, J. D. (2001). ‘The biological basis of speech: what to infer from talking to the animals’, Psychological Review, 108, 523–549.

Vernes, S. C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J. M., & Davies, K. E, Geschwind, D. H., & Fisher, S. E. (2007). ‘High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders’, *American Journal of Human Genetics*, 81, 1232-1250.

Vernes, S. C., Newbury, D. F., Abrahams B. S., Winchester, L., Nicod, J., Groszer, M., Alarcon, M., Oliver, P. L., Davies, K. E., Geschwind, D. H., Monaco, M. D., Fisher, E. F. (2008). ‘A functional genetic link between distinct developmental language disorders’, *The New England Journal of Medicine*, 59, 2337-2345.

Vernes, S.C., Oliver, PL, Spiteri, E, Lockstone, H.E., Puliyadi, R., Taylor, J. M., Ho, J, Mombereau C., Brewer, A., Lowy, E., Nicod, J., Groszer, M., Baban, D., Sahgal N., Cazier, J.B., Ragoussis, J., Davies, K.E., Geschwind, D.H., Fisher, S.E. (2011) 'Foxp2 Regulates Gene Networks Implicated in Neurite Outgrowth in the Developing Brain.' PLoS Genet 7(7): e1002145.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). 'Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech', *Journal of Cognitive Neuroscience*, 13, 994–1005.

Wang, B., Lin, D., Li, C., & Tucker, P. (2003). 'Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors', *Journal of Biological Chemistry*, 278, 24259-24268.

Wanner, E. & Gleitman, L. R. (1982). *Language Acquisition: The State of the Art.* Cambridge: Cambridge University Press.

Watkins, K. E., Dronkers, N. F., Vargha-Khadem, F. (2002). 'Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia', *Brain*, 125, 452-64.

Wheelwright, N. T., Swett, M. B., Levin, I. I., Kroodsma, D. E., Freeman-Gallant, C. R., & Williams, H. (2008). 'The influence of different tutor types on song learning in a natural bird population', *Animal Behaviour*, 75, 1479-1483.

White, S.A, Fisher, S.E, Geschwind, D.H., Scharff, C., Holy, T.E. (2006) 'Singing mice, songbirds, and more: models for FOXP2 function and dysfunction in human speech and language.' *J Neurosci.* 26, 10376-9.

Williams, H. & Nottebohm, F. (1985). 'Auditory responses in avian vocal motor neurons: a motor theory for song perception in birds', *Science,* 229, 279–282.

Williams H, Staples K. (1992) Syllable chunking in zebra finch (Taeniopygia guttata) song. *J Comp Psychol.*106, 278-86.

Wohlgemuth, M. J., Sober, S. J., & Brainard, M. S. (2010). 'Linked control of syllable sequence and phonology in birdsong', *Journal of Neuroscience*. 29, 12936–12949.

Yu, F., Keinan, A., Chen, H., Ferland, R. J., Hill, R. S., Mignault, A. A.,Walsh, C. A. & Reich, D. (2009) Detecting natural selection by empirical comparison to random regions of the genome. *Hum. Mol. Genet.* 18, 4853–4867. (doi:10.1093/hmg/ddp457)

Zhang, J., Webb, D. M., & Podlaha, O. (2002). 'Accelerated protein evolution and origins of humanspecific features: Foxp2 as an example', *Genetics*, 162, 1825-1835.

Zirin, R. A. (1980). 'Aristotle's Biology of Language', *Transactions of American Philological Association*, 110, 325-347.