Title
Possible Precursors of Syntactic Components in Other Species

ISBN
978-0-262-01356-7

Authors
White, SA
Hilliard, HT

Publication Date
2009

Peer reviewed

8

Possible Precursors of Syntactic Components in Other Species

Austin Hilliard and Stephanie White

Abstract

Please provide.

Introduction

Human linguistic syntax is the system of forming complex signals and the mapping of these signals onto conceptual/intentional representations. As such, syntax provides compositionality: it serves to combine a finite number of meaningful units to produce an infinite variety of sequences with larger meanings. This ability for limitless recombination appears uniquely human (Hauser et al. 2002). However, less sophisticated rule systems for recombining units and recognizing sequences of communication signals, or other stimuli, exist in animals and form the focus of this contribution.

Why Other Species?

How the human language singularity evolved remains a puzzle. We argue that a comparative approach examining animal communication and its underlying neural basis can inform studies of human language evolution. Biologists refer to the study of animals for the purpose of understanding the human condition as using an "animal model." While no animal model can fully capture any aspect of language, preadaptations for subcomponents of language, including syntax, exist most likely in nonhuman species. The alternative idea that syntax appeared *in toto*, via one fortuitous mutation, in a manner unique to humans, is not parsimonious. Rather, similar environmental and biological constraints on vocal communication have likely driven parallel solutions across multiple animal groups, examples of which we discuss below, with an emphasis on the meaningless syntax of songbirds. These preadaptations, coupled with the

biological and cultural evolutionary interactions, discussed by Számadó et al. (this volume), may have combined in the hominid lineage to produce the unique language phenotype.

Any user of vocal communication signals must be able physically to transmit and receive these sounds. A co-requisite is the ability to convey (even passively, as in advertising the size of one's larynx simply by vocalizing) and decode behaviorally important information in vocal signals. Here, we focus on a narrower target, namely, vocal learners, or animals with the experience-dependent capacity to learn these production skills, and the neural basis that gives rise to this capacity. Vocal learning requires the ability to coordinate precisely and rapidly complex sequential movements of lingual, vocal, and respiratory musculature in order to mimic conspecifics or create new sounds. Although innate abilities may point to parallel mechanisms, from a biological perspective, neural circuitry that is capable of learning seems most likely to capture the unbounded features of language, a learned behavior. We sometimes broaden our perspective from the learning of vocal skills to the learning of any motor skill. The mechanisms which underlie vocal learning in more common neural circuits may form the evolutionary roots for the neural basis of speech.

Which Other Species?

Vocal learning depends upon hearing conspecific vocalizations as well as one's own. Comparison of these inputs determines whether neurobiological changes must occur for adaptive modifications of vocal output. Tests for the vocal learning capacity often rely on deprivation of these acoustic inputs and determination of whether the subsequent vocal output is abnormal. Deprivation can be drastic (such as deafening), dramatic (as in rearing in the absence of conspecific vocalizations), or refined (e.g., transient distortion of key auditory inputs). A noninvasive method examines whether changes in vocal output during normal development are more substantial than those expected due to physical maturation of the vocal apparatus (Fitch 1997) or are uncharacteristic of their species. Marine mammals, bats, and elephants are thereby considered vocal learners (Boughman 1998; Janik et al. 2006; Suzuki et al. 2006; Poole et al. 2005). By a majority of these tests, passerine birds of the oscine suborder, known as songbirds, are vocal learners as well.

Birdsong as a Unique Model System: Parallels between Song and Speech

About half (~4,500) of the extant avian species are songbirds; their phenotypic distinction is that they learn part of their vocal repertoire. This ability is shared with hummingbirds and parrots, which are in separate avian orders, raising the hypothesis that vocal learning emerged three times, independently, in the avian lineage. Among the identified vocal learners, songbirds are the most amenable to controlled experimentation. Certain species, such as white-crowned sparrows

(*Zonotrichia leucophrys*), zebra finches (*Taeniopygia guttata*), and Bengalese finches (*Lonchura domestica*, also known as society finches), are small and breed in the laboratory. As a result, much about their song learning and its underlying neural bases is known. Songbird researchers divide song into songs, bouts, phrases, motifs, syllables, and notes (Konishi and Nottebohm 1969). Notes are the smallest unit, combining together to form syllables. Two or more syllables may group together to form a phrase. A motif is a sequence of notes and/or syllables that are repeated in a stereotyped order. One or more motifs or phrases followed by an interval of silence constitute a bout of song. (Brenowitz et al. 1997). Song phonation refers to the acoustic features such as amplitude, mean frequency, frequency modulation, amplitude modulation, and entropy. Song syntax refers to the temporal order of these features within a song (e.g., the order of syllables within a motif or phrases within a song).

Song and speech learning share key features, including dependence on hearing and on social interactions with conspecifics (Doupe and Kuhl 1999). Both occur during critical developmental phases beginning with an early perceptual phase where, in the case of songbirds, the song of an adult male is memorized. In humans, this corresponds to a time of universal speech perception when a baby listens to speech, but does not yet produce any learned vocal output. In a second phase, known as sensorimotor learning, both songbirds and humans practice and refine their own vocalizations in order to mimic adult sounds. In normal children, the onset of sensorimotor learning is marked by babbling at ~6 months and results in first words at ~one year, with two word combinations occurring at ~two years. Thereafter, word production and syntactical recombination takes off and can continue, to a degree, throughout life. In contrast, the degree of flexibility in mature songbirds depends upon the species.

By the time zebra finches reach sexual maturity, their previously variable songs have stabilized through a process termed *crystallization*. The unchanging song of adult zebra finches appears to contrast with the less limited capacity of human language. Yet other songbird species, such as mockingbirds, are capable of learning new songs throughout life. The maintenance of mature zebra finch song requires continuous auditory feedback, as it gradually deteriorates in deafened birds (Nordeen and Nordeen 1992; Brainard and Doupe 2000). Deafening-induced song deterioration is even faster in Bengalese finches (Woolley and Rubel 1997). Adult birdsong can be temporarily disrupted in intact birds exposed to abnormal auditory feedback (Cynx and Von Rad 2001). Similarly, adult speech depends on ongoing auditory feedback. Finally, as is evident to anyone trying to learn a new language after puberty, the human faculty for language peaks in youngsters. Thus, comparison of all of birdsong to human speech reveals both developmental constraints as well as relative openness to experiential input throughout life, and ongoing dependence on audition.

Beyond behavior, the neuroanatomical circuits underlying song and speech show additional parallels (Jarvis 2004). While the lateralized cortical regions classically known as Broca's and Wernicke's areas are arguably unique to

humans (cf. Taglialatela et al. 2008 for chimpanzees), emerging evidence implicates subcortical brain structures, notably, the cerebellum and basal ganglia as critical for language and speech. The planning and execution of complex motor skills involves circuits that run through the cortex, basal ganglia, thalamus, and back to the cortex (Liebermann 2006). Relevant here, songbirds use similar loops during learning and production of song (Figure 8.1).

The collection of brain areas specialized for song is referred to as the song circuit and is well characterized, partly due to its sexual dimorphism in species such as the zebra finch (Nottebohm and Arnold 1976). Subregions within the cortical-like pallium, the basal ganglia, and thalamus are prominent and interconnected only in males. Neurons within these subregions are dedicated to song, lacking regular firing patterns during performance or perception of other behaviors or stimuli. The song circuit is comprised of two interconnected pathways that each stem from the cortical-like (pallial) HVC, analogous to association cortex in humans (Figure 8.1). In the first, known as the vocal motor pathway, auditory inputs enter the circuit at HVC. A subset of HVC neurons projects their axons to a region analogous to primary motor cortex. These neurons, in turn, make direct projections onto brainstem motor neurons, which control the muscles used in singing (Nottebohm et al. 1982; Wild 1993). The direct connectivity between high-level cortical-like regions and the motor neurons that control the song organ is reminiscent of similar motor cortical projections in humans (and raccoons), which bypass brainstem way stations and directly contact finger motor neurons (Sereno 2005). This privileged access between high-level brain area and effector may underlie the enormous flexibility of motor output in each system. The second pathway indirectly links HVC to RA via forebrain structures and includes the striatal nucleus Area X, and the lateral portion of the magnocellular nucleus of the anterior nidopallium

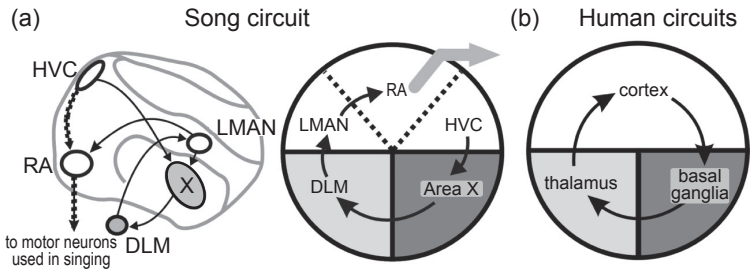

Figure 8.1 Schematic comparison of the avian song circuit and human cortico-basal ganglia-thalamo-cortical circuitry. The cortex is white, basal ganglia dark gray, and thalamus, light gray. (a) A composite sagittal view of songbird telencephalon is depicted on the left. Auditory input (not shown) enters the song circuit at HVC, the neurons of which contribute to two pathways: the vocal motor pathway (stippled arrows) and the anterior forebrain pathway (plain arrows; adapted from Teramitsu et al. 2004). (b) Human circuits. HVC: high vocal center; RA: robust nucleus of the archistriatum; DLM: dorsolateral thalamus; LMAN: lateral magnocellular nucleus of the anterior nidopallium.

(LMAN). LMAN neurons rejoin the pathways by projecting to RA. This anterior forebrain pathway (AFP) is required for song modification, most prominent during sensorimotor learning (Bottjer et al. 1984; Scharff and Nottebohm 1991; Okuhata and Saito 1987; Sohrabji et al. 1990), but also during maintenance of adult song (Brainard and Doupe 2000; Williams and Mehta 1999). The essential point is that the pathway allowing for song modification and maintenance is similar to the cortical loops that underlie the planning and execution of complex sequential movements in humans.

The FOXP2 Puzzle Piece as a Genetic Example

A transcription factor known as FOXP2 was identified in 2001 as the monogenetic locus of a mutation causing an inherited speech and language disorder (Lai et al. 2001). Instantly, FOXP2[1] flashed as a key piece in the puzzle that, through interactions with other molecules, patterns the brain for language. Although the inherited disorder is not chiefly syntactical in nature, there is a morphosyntactic component (see below). Further, the FOXP2 story demonstrates how biologists, neurologists, and linguists can work collaboratively to uncover pieces of the neuromolecular puzzle underlying language.

The KE Family and Others

The FOXP2 discovery emerged from clinical characterization of the KE family. Of several expressive and receptive linguistic features, their most prominent deficits are in sequencing of orofacial movements especially those required for speech, referred to as verbal dyspraxia. Whether their cognitive deficits are in addition to, or a consequence of, their motor deficits is an area of ongoing investigation. Affected members have morphosyntactic difficulties as exemplified by their inconsistency in adding suffixes such as *s* for plurals or *-ed* for actions occurring in the past (Vargha-Khadem et al. 2005). Moreover, their gross orofacial dyspraxia is relevant here, given our broad interest in the capacity to generate complex sequenced movements.

The identification of the genetic basis for the KE phenotype hinged on the discovery of an unrelated individual, CS, who exhibited similar deficits. Examination of CS's chromosomes revealed an easily detectable rearrangement, one end of which interrupted the gene encoding FOXP2, a transcription factor with unknown neural function. FOXP2 was known to be a transcriptional repressor in lung, binding to sequences in the noncoding region of its target genes, thereby decreasing their expression levels. Subsequent analysis of the KE family FOXP2 sequence revealed a single point mutation in the DNA binding domain

[1] By convention, human "FOXP2" is capitalized, mouse "Foxp2" is not, and "FoxP2" denotes the molecule in mixed groups of animals. Italics are used when referring to genetic material such as *FoxP2* mRNA (Carlsson and Mahlapuu 2002).

of the molecule. Since 2001, additional cases of verbal dyspraxia linked to *FOXP2* have emerged, including a mutation that truncates the protein prior to the DNA binding domain (Macdermot 2005). These mutations impede FOXP2 from binding to its transcriptional targets (Vernes et al. 2006).

Imaging studies comparing affected to unaffected KE family members have revealed the neuroanatomical bases for their deficits. Affected individuals have bilateral abnormalities in the basal ganglia and cerebellum, in addition to cortical abnormalities including the Broca's area in the inferior frontal gyrus. Altered amounts of gray matter in these regions are accompanied by their underactivation during tasks of verbal fluency (Vargha et al. 2005; Belton et al. 2003; Liegeois et al. 2003). These findings suggest that a mutant copy of *FOXP2* during development results in the malformation of brain structures later used in the control of orofacial musculature important for speech.

Progress in Animal Models

Advances in understanding FoxP2 neural function have been made in songbirds and mice (White et al. 2006). As in humans, FoxP2 is expressed in the cortex/pallium, striatum, and thalamus of these animals during development, consistent with a role in forming these structures (Ferland et al. 2003; Lai et al. 2003; Takahashi et al. 2003; Teramitsu et al. 2004; Haesler et al. 2004). In songbirds, expression persists in adults when *FoxP2* mRNA and protein are actively regulated in the striatal song circuit region Area X when birds sing under certain social conditions (Teramitsu and White 2006; Miller et al. 2008). This implicates the molecule in the functional use of song circuitry. Beyond correlation, tests of molecular function rely on being able to manipulate molecular expression. The technology for manipulating genes in birds is not yet routine. Thus, Haesler et al. (2007) exploited viruses for their ability to enter cells and express foreign molecules. They used lentiviruses to express short hairpin sequences of RNA designed to knockdown FoxP2 expression. Virus was injected bilaterally into Area X of juvenile birds, which were then given the opportunity to learn their tutors' songs. Strikingly, in adulthood, their song copies lacked precision. The acoustic abnormalities were reported to resemble those exhibited by children with developmental verbal dyspraxia in that syllable structures and duration were abnormally variable. Regarding syntax, the jury is still out. No difference in the consistency with which knockdown birds ordered their syllables was detected, although the knockdown group had greater variability on this measure. Further, in the three sets of exemplar spectrograms, only the control pupils' syllables occur in the same order as those of their tutors. Perhaps more robust alterations in syntax would emerge with testing of more subjects, or with a consistently high percent of viral transduction (20 ± 10% of the Area X neurons were affected), or by testing other species with more sophisticated syntax e.g., the Bengalese finch).

The technology for introducing inherited transgenes is well developed in mice, enabling functional tests of molecular manipulations. Thus far, four groups have used transgenics to alter Foxp2 in mice. In the first, Shu et al. (2005) produced a null mutation, creating intermediate Foxp2 levels in heterozygotes and undetectable levels in homozygotes compared to wild-type animals. As expected, the homozygous phenotype was the most dramatic with pups dying by postnatal day 21. Several behavioral tests were run, including examination of ultrasonic vocalizations. Although no spatial learning deficit was detected with the Morris water maze test, a major finding was that stranded heterozygous pups had reduced numbers of ultrasonic isolation calls, which typically cause the dam to retrieve the pup. The calls were reported to be acoustically normal, yet their decreased numbers in heterozygotes relative to wild-type pups was taken as evidence for a specific effect of Foxp2 on mouse vocalizations.

Subsequent studies did not aim to knock out Foxp2 but rather to mutate it. Groszer et al. (2008) generated lines of mutant mice by exposing founder males to N-ethyl-N-nitrosourea (ENU) which induces mutations randomly across the genome. The genomic DNA from >5,000 offspring was then screened. Incredibly, a mouse that carried the KE-like mutation in its Foxp2 gene was identified. While labor intensive, the ENU methodology does not introduce stretches of foreign DNA into the host, and backcrossing is used to remove nonrelevant mutations. In contrast to the heterozygous nulls described above, the heterozygote pups generated by Groszer et al. emitted similar numbers of normally structured isolation calls as their wild-type littermates. Analysis of different calls and call properties led the authors to conclude that severe Foxp2 mutations cause developmental delays. Rather than loss of a specific function of Foxp2, these generalized delays may underlie the lower number of calls in the heterozygous nulls. Assuming such calls are unlearned (which is likely given their function in newborn mice), the normal calls and call numbers of the KE-like heterozygotes are not surprising, as they reinforce the notion that FoxP2 is critical for learned skills such as speech in humans or other procedurally learned behaviors in nonvocal learners.

In line with this idea, the heterozygote KE-like mice showed deficits on two assays of motor skill learning: the accelerating rotorod and the tilted voluntary running wheel. Further, in these mice, neurons in the dorsal striatum, a region implicated in motor skill learning (Dang et al. 2006), lacked a form of synaptic plasticity known as long-term depression. Synapses in the cerebellar cortex also showed altered plasticity. Together, these findings depict a developmental role for Foxp2 in motor-skill learning, involving cortico-striatal and cortico-cerebellar circuitries, that extends to the learned speech of humans, but not to the unlearned calls of mice. Fujita et al. (2008), however, used targeted genetic recombination to generate KE-like heterozygote pups that had disrupted calls, raising a discrepancy between studies. This method necessarily leaves foreign genetic material in the animal, which may account for the different findings. Some resolution may emerge as other mice are generated with “conditional”

mutations that can be switched on or off throughout life. Interestingly, while all three studies discussed thus far detected cerebellar abnormalities, only Groszer et al. (2008) found striatal ones.

Finally, rather than creating a loss of function phenotype, Enard and colleagues investigated whether the two amino acids from the FOXP2 sequence that, among primates, are unique to the human form (Asn at position 303 and serine at 325, Enard et al. 2002), would promote a new function when inserted into mice. At the time of this writing, the full results are not yet available, but a preliminary report suggests that the human-like sequence promotes neurite outgrowth and synaptic connectivity in the striatum (Enard et al. 2002). This observation, together with the disrupted striatal plasticity in mice with the KE-like mutation, the imprecise song copying of birds with lower striatal FoxP2 levels, and the altered striatal structure and activation in affected KE family members reinforce the importance of the striatum in skill learning. Perhaps FoxP2 function is critical for species-typical skills, be it mouse locomotor coordination, birdsong, or human speech.

FoxP2: Achieving Specificity

FoxP2 is expressed in most, if not all, body organs (Lu et al. 2002) and is present across the animal kingdom. The human mutant phenotype, however, appears restricted to neural function and, most generally, to learned motor control of the mouth and face, which does not extend to other body parts (Vargha-Khadem et al. 2005). How this specificity occurs requires consideration not just of the sequence that makes up FoxP2 itself, but also of the gene regulatory networks in which it participates.

FoxP2 Molecular Evolution

It is tempting to hypothesize that in certain phylogenetic lineages the molecular sequence of FoxP2 is critical for specialized acoustic motor function. However, only among primates is there an obvious correspondence between amino acid changes and vocal learning capacity. Accelerated evolution occurred in primates along the lineage between chimpanzees and humans (Enard et al. 2002), and in the human coding sequence, two amino acids are distinct (Zhang et al. 2002). Carnivores (who are not thought to learn vocalizations) share one of these two. Further confusing the picture, there is no correspondence between FoxP2 sequences and vocal learning abilities in birds (Webb and Zhang 2005) or acoustic motor abilities in bats (Li et al. 2007). Despite ignorance of the pressures driving accelerated evolution in primates, the finding of the modern human FOXP2 form from Neanderthal bones reopened the question of whether these archaic hominids possessed a protolanguage (Krause et al. 2007).

The difficulty in relating a single gene sequence to a complex behavioral phenotype highlights the fact that genes do not specify behaviors or cognitive

processes (Fisher 2006). Instead, they make bits of biological machinery, things like signaling molecules, receptors, and regulatory factors such as FoxP2. These tiny machines interact in complex networks, and those networks are themselves cogs in larger hierarchical arrangements (Barabasi and Oltvai 2004). Thus, although *FOXP2* is implicated in speech and language impairment and neuroligins *NLG3* and *NLG4* in autism (Jamain et al. 2003), these genes cannot be considered "a gene for speech and language" or "genes for sociality." Coding variants of *FOXP2* have not been found in specific language impairment (SLI), and the relevance of *NLG3* and *NLG4* variants for common disorders remains open (Fisher 2006). The information carried in the genotype is simply the starting point for cascades of molecular, cellular, and systems-level interactions in the brain. As such, the discovery of the KE family and others with similar mutations is a fortuitous clue to the puzzle of the complex genetic basis of speech and language. The molecular targets of FoxP2's transcriptional regulation are the next set of interactors in the puzzle.

Gene Targets

Recently, the technique of chromatin immunoprecipitation coupled with gene promoter microarrays (ChIP-chip) was used by two groups to identify molecular targets of FOXP2 in human tissue (Spiteri et al. 2007; Vernes et al. 2007). This method uses an antibody to grab onto FOXP2 *in flagrante* (i.e., within a biological tissue and in the act of binding to the noncoding promoter regions of its target genes). After chemical dissociation, these DNAs are hybridized to microarray chips spotted with many samples of different known genes. A positive hybridization signal indicates that the spotted gene is likely a FOXP2 target. A variation of this method is referred to as ChIP-seq, in which, following dissociation, the target DNAs are directly sequenced to reveal their identities, an approach that we will return to at the end of this section.

Using ChIP-chip, Vernes et al. (2007) looked for targets in human SH-SY5Y cells, useful for studying neural processes while Spiteri et al. (2007) investigated FOXP2 targets in human fetal basal ganglia (BG) and inferior frontal cortex (IFC)—two main areas of dysfunction in people with FOXP2 mutations—and in human fetal lung. Spiteri et al. identified ~175 targets in each tissue, many of them overlapping. Most intriguing are the eight targets enriched in both brain areas but not in lung, three of which are involved in central nervous system development, including cortical patterning. Gene ontology and pathway analyses were used by both groups to investigate the strongest signals and revealed enrichment for molecules involved in ion transport, cell signaling, neurite outgrowth, and synaptic transmission. The Wnt/Notch signaling pathway was highlighted in the SH-SY5Y cells (Vernes et al. 2007). In mammals, *Wnt* genes are known to play a major role in forebrain patterning during development, consistent with the idea that FOXP2 targets in this pathway mediate structural changes in these regions. Additional targets have links to learning

(Spiteri et al. 2007), which suggests a function for FOXP2 signaling cascades in activity-based sculpting of neural connections. This is intriguing in light of our own work, which shows down-regulation of FoxP2 in the striatal Area X of the song circuit when adult birds practice, but not when they perform their songs (Teramitsu and White 2006). As song practice is more acoustically variable, we posit a role for avian FoxP2 in adult neural and behavioral plasticity. Fourteen of the FOXP2 targets found by Spiteri et al. (2007) show accelerated evolution likely via positive selection, together comprising a genetic cohort potentially related to human cognitive specializations integrated by the BG and IFC, including speech and language. The significant overlap between the targets found in the BG and IFC and the targets identified in SH-SY5Y cells further validates both studies.

Significance of Gene Networks

Uncovering the genetic basis for a complex trait is a task rife with many challenges, yet just as many adaptive strategies are coming into use (Fisher et al. 2003; Felsenfeld 2002). One exciting strategy for investigating gene networks underlying complex traits was developed at our home institution, UCLA. Weighted gene co-expression network analysis (WGCNA) identifies groups of functionally related genes through statistical analysis of gene expression microarray data (Zhang and Horvath 2005). Until recently, the interpretation and analysis of microarray data has resulted in little more than lists of potentially interesting genes, left to be functionally validated via more traditional molecular techniques. That is because researchers have generally only considered genes in isolation, and whether they are expressed at different levels between control and experimental samples. A major drawback of this approach is the statistical difficulty inherent in comparing long lists of data points from insufficiently large sample sizes.

WGCNA takes a step back from the search for single differentially expressed genes, instead identifying groups of genes whose expression levels co-vary, called modules. The modules are defined via hierarchical clustering and are correlated to the trait of interest. In this way, multiple hypothesis problems are alleviated, and the investigator can be confident in the reproducibility of his/her results. By considering the functional activity of groups of genes, WGCNA has thus far yielded robust and compelling findings for various traits, including identification of novel genes involved in brain cancer, gene pathways relevant to atherosclerosis, and genes that are key drivers of evolutionary change in humans (Horvath et al. 2006; Gargalovic et al. 2006; Oldham et al. 2006). Language may very well be the most complex of traits. Application of WGCNA or similar strategies to high throughput data on FOXP2 gene targets promises to help piece together the molecular puzzle underlying language.

The CNTNAP2 Connection

Discovery of gene networks using microarrays is technically limited by the number of genes that are printed on the arrays; often an incomplete set. An alternative approach is to identify target genes by direct, so-called "shotgun" sequencing. This unbiased approach was recently used by Vernes et al. (2008) to great effect. It identified the contactin-associated protein-like 2 gene (*CNTNAP2*) that encodes a neurexin protein, called CASPR2, as a direct target of FOXP2 transcriptional repression. In the brain, association of presynaptic neurexins with postsynaptic neuroligins is thought to be a key event in synaptogenesis (O'Connor et al. 1993; Dean et al. 2003). Accordingly, during human fetal brain development, *CNTNAP2* expression is enriched in areas of the cortex that give rise to language, while in developing rodents, its cortical expression is diffuse (Alarcón et al. 2008). The human *CNTNAP2* cortical pattern is opposite that of *FOXP2* (i.e., *FOXP2* levels are high where *CNTNAP2* levels are low), consistent with FOXP2 repression of this transcript. As mentioned above, FOXP2 variants have not been associated with SLI nor with common developmental disorders in which language is delayed or impaired, such as in autism. In sharp contrast, certain *CNTNAP2* variants in autistic children are associated with the age at first word (Alarcón et al. 2008). Further, Vernes et al. (2008) have found that genetic polymorphisms of *CNTNAP2* in children with SLI are correlated with their ability to do a nonword repetition task. The viewpoint that *FOXP2* is not a gene for language, but rather a key molecular piece that connects with many others in the language puzzle, is now directly supported by its transcriptional repression of *CNTNAP2*. This interaction connects FOXP2 with CNTNAP2-related language disorders in which FOXP2 was not otherwise implicated.

Syntax in Birdsong

Turning from the discussion on FoxP2, we now review what is known about "song syntax" in songbirds. In the following sections, we discuss birdsong syntax, its dependence on critical developmental phases for learning and auditory feedback, its role in perception, its social regulation, and its neural coding.

Phonetical Syntax

Birdsong syntax is defined simply as the temporal sequence in which discrete units of song (notes, syllables, phrases, motifs, bouts) are produced. These units do not carry independent meaning, and thus bird syntax has been called "phonological syntax" by analogy with the organization of independently meaningless phonemes into morphemes or words in language (Marler 1977). Since morphemes are symbolically meaningful, however, the analogy is not perfect.

While birdsong can communicate species or individual identity and advertise mating or territorial ownership (Doupe and Kuhl 1999), it is not compositional; the position of a single note within a syllable, a single syllable within a motif, or a single motif within a bout does not by itself provide new semantic information. Thus, we adopt the terms "phonetical syntax" to refer purely to the sequence of sounds. Whatever vocabulary we use, in both birdsong and language, discrete acoustic units are produced and then combined according to specific sets of learned rules. In both, fine motor sequences are executed using the vocal motor apparatus, a task dependent on subcortical structures such as the basal ganglia and cerebellum.

Development

Similar to human speech and language learning, all aspects of birdsong development show critical constraints on acquisition as well as dependence on auditory feedback. In juvenile white-crowned sparrows syntactical and phonetical (i.e., the spectral features of song syllables) cues guide selective song learning during development (Soha and Marler 2001) Behavioral studies of zebra finch song development indicate that learning the spectral content of tutor notes and syllables occurs before learning how to order them in time (Tchernichovski et al. 2001). As development advances, zebra finches produce a greater variety of syllables delivered in sequence, reminiscent of the emergence in human infants of reduplicated babbling and the transition to variegated babbling. Multiple studies suggest that different neural pathways underlie the mimicry of individual syllables versus syllable order (Vu et al. 1994; Yu and Margoliash 1996; Hahnloser et al. 2002), and that aspects of song syntax can be acquired independently of spectral content (Helekar et al. 2003).

In 2003, Funabiki and Konishi used sustained white noise exposure to deprive zebra finches of auditory feedback during the sensorimotor phase of song development. Birds were reared with their biological parents until the onset of sensorimotor learning. Experimental birds were then moved into sound attenuation chambers with continuous loud noise playing in the chamber, depriving the birds of any auditory feedback. They were kept in noise for ~1–6 months before being released and allowed to sing under normal acoustic conditions. They learned phonetical features of the tutor song just as well as birds that did not endure noise exposure, with similar developmental trajectories. Interestingly though, only those birds released from noise before 80 days of age were able to reproduce correctly the syntax of their tutor. All birds released after 80 days were unable to learn song syntax, perhaps indicating stronger developmental constraints on syntax than on phonation.

To examine the consequences of perturbing auditory feedback in real time, Sakata and Brainard (2006) interleaved trials of perturbed versus normal feedback to Bengalese finches during singing. They showed that, as for human speech, abnormal feedback disrupts both the sequencing and timing of song

within tens of milliseconds. Specifically, altering auditory feedback reduced stereotypy in syllable transition probabilities and could elicit the production of completely novel sequences of syllables during what were otherwise completely stereotyped sequences.

Perception

Syntax is known to play a role in song perception, as evidenced by multiple studies using operant conditioning paradigms to train songbirds to discriminate between songs. Comparing results from different studies, Bengalese finches weigh syntax more heavily than zebra finches when doing discrimination tasks (Braaten et al. 2006), perhaps reflecting a functional importance of higher syntax variability in Bengalese finches. Gentner and colleagues used operant conditioning in European starlings (*Sturnus vulgaris*) to investigate the perceptual mechanisms underlying the vocal recognition of individual conspecifics (Gentner and Hulse 1998). After training, birds were able to discriminate individuals on the basis of novel song bouts, mediated by the memorization of specific motifs and the sequential ordering of motifs within bouts. A later expansion upon this study implicated HVC in the process of forming learned associations among conspecifics (Gentner et al. 2000) to show that HVC is important for the production of song syntax as well. The dual roles of HVC in song perception and production place it in an intriguing category of brain regions with "mirror like" properties, a topic that we will return to later.

Again using operant conditioning, starlings were trained to classify subsets of motif sequences in two different types of artificial starling "languages" (Gentner et al. 2006). Different types of typical starling motifs were used to generate a context-free grammar (CFG; A^nB^n) which entailed recursive center-embedding (i.e., a structure of one type, *AB*, is embedded in another instance of itself *AABB*) and a finite-state grammar (FSG: $(AB)^n$). After thousands of training trials, nine of eleven birds were able to classify the CFG and FSG sequences accurately, as well as successfully classify novel sequences generated using the same rules. This study followed a similar experiment testing the syntactic capabilities of cotton-topped tamarin monkeys (*Saguinus oedipus*) wherein the authors concluded that the tamarins were able to master a FSG ($(AB)^n$), but not a CFG (A^nB^n) (Fitch and Hauser 2004). Humans trained on the same stimuli did master the CFG. This seems to substantiate an earlier claim (Hauser et al. 2002) that the capacity for recursion is unique to humans, but the questionable ecological relevance of the stimuli (recordings of human male and female voices) and small size of the training sets (only 2 samples per set) used in the tamarin study have prompted criticism. A major problem is the legitimacy, or lack thereof, of making conclusions based on generalizing from such small training sets to entire grammars.

The stimuli in the starling study were actual starling song motifs ("rattle" motifs were *A*, and "warble" motifs were *B*) and were varied such that multiple

kinds of rattle and warble motifs were used, resulting in larger sets of ecologically valid stimuli. A series of probe tests provide compelling evidence that the successful starlings were not merely approximating the recursive structure of the CFG by learning an equivalent FSG, or using alternate discrimination strategies such as attending to only primary or terminal patterns or simply counting *A/B* transitions. This suggests that starlings can recognize strings formed using a recursive center-embedding rule and challenges the claim that only humans are capable of learning grammars with such a rule. If this is so, we now have at our disposal a well-established model system within which we can probe the neural systems encoding a learned CFG. However, it is important to keep in mind the immense amount of training required by the successful starlings (~9,400 to ~56,000 trials) and consider that even humans can fail to learn center-embedding grammar in artificial languages lacking semantic content, as shown by Perruchet and Roy (2005). Note, however, a general problem with these experiment that animals in fact could solve the problem by simple counting (Corballis 2007b).

If even humans can fail to learn a CFG, where does this leave us? Note that the amount of center-embedding in the stimuli presented to Perruchet and and Rey's subjects is no greater than that present in the simple sentence *Either they learned or they did not*. Hurford contends that these human subjects fared so poorly due to a lack of semantics in the artificial language (Hurford 2009). Perhaps they may have been successful if they had been sufficiently motivated to undergo exhaustive training like that endured by the starlings. Perhaps the difference between human and animal syntax can be accounted for by the lack of semantic interpretation in animal syntax coupled with stricter computational restraints in animal brains and is not so much a matter of FSGs and CFGs. It is possible that humans and animals all use some type of CFG, but humans simply exploit CFG to a greater depth, ultimately taking advantage of mechanisms such as recursion thanks to more formidable neural processing power combined with well-developed symbolic representational abilities. Clearly, birdsong lacks semantics, but if starlings or finches could map syllables or motifs of song onto conceptual or intentional representations of the world, perhaps a relatively simple (due to weaker neural processing power) compositional syntax would emerge in birds as well.

Social Context

Social regulation of syntax variability has been described in adult Bengalese finches (Sakata et al. 2008). Songbirds can sing in one of at least two social contexts: (a) directing song to a conspecific female, known as directed (FD) song, or (b) singing alone, known as undirected (UD) song. Bengalese finches were used to test specifically for social modulation of song syntax. Unlike zebra finch song, Bengalese song retains moment to moment variability in syllable sequencing (Okanoya 2004a). In zebra finches there is greater variability

in syllable structure (mostly driven by variability in fundamental frequency) during UD song when compared to FD song (Kao et al. 2005). Analogous modulation was found in the syntax of Bengalese finch song, with greater variability in sequencing during UD over FD songs (Sakata et al. 2008).

Interestingly, social context regulation has been demonstrated at both the neural systems and molecular levels (Jarvis et al. 1998; Hessler and Doupe 1999; Hara et al. 2007). Acute down-regulation of *FoxP2* has been observed in Area X of zebra finches singing UD song when compared to FD song, indicating a possible role for FoxP2 in modulating behavioral variability (Teramitsu and White 2006).

Social context modulates dopamine (DA) levels in Area X, with levels being higher in FD song (Sasaki et al. 2006). The DA system is important for motivation and is thus a sensible candidate for driving social context modulation in the song circuit. In line with this notion, Sakata and Brainard (2006) propose a model of quick acting dopaminergic influence over the social modulation of song syntax variability found in Bengalese finches. The changes they observed occurred on a small enough time scale (1–3 minutes) which effectively rules out slower acting transcription factor-mediated processes as a biological mechanism. However, the medium spiny neurons expressing FoxP2 in Area X receive strong dopaminergic input. Thus the functional relationship between *FoxP2* and its gene targets, striatal microcircuits, DA, and social context regulation of song features like syntax variability requires further study.

Coding

The neural coding of syllable sequence production by the vocal motor pathway has been examined by making electrophysiological recordings from neurons in RA (analogous to primary motor cortex in mammals; projection neurons here directly innervate syringeal motor neurons) and HVC (analogous to premotor cortex in mammals, and presynaptic to RA). During singing, RA neurons generate a complex sequence of high-frequency spike bursts, reproduced precisely each time the bird sings a motif (Yu and Margoliash 1996). HVC neurons come in three basic types: interneurons (HVC_I), neurons projecting to Area X (HVC_X), and neurons projecting to RA (HVC_{RA}) (Hahnloser et al. 2002). By recording from HVC_{RA} neurons during singing, Hahnloser et al. showed that this specific subpopulation generates sparse patterns of bursts that are time-locked to the song, with each HVC_{RA} neuron emitting only one burst of spikes per song motif. The timing is extremely precise, with a jitter of less than 1 ms relative to the song. Recording simultaneously from HVC_{RA} and RA neurons in sleeping birds revealed that RA neurons generate complex burst sequences temporally locked to the sparse sequences in HVC as well as HVC_{RA} population activity correlated with bursting in RA. Thus, it is likely that sparsely firing HVC projection neurons are driving bursting activity in RA via direct feedforward input (further discussed below in the section on Modeling).

Population activity of HVC_{RA} neurons has since been highly correlated with syllable sequencing (Kozhevnikov and Fee 2007) Also, there is no correlation between the firing of HVC_{RA} neurons and spectral features of song syllables, or between the timescales of vocal dynamics and neural dynamics in HVC. This suggests that HVC codes for the temporal order of syllables in song and not for spectral structure, supporting the hypothesis of separate neural substrates underlying phonation and syntax in birds.

Quantification and Comparison

Birdsong syntax is typically analyzed in one of two subtly distinct ways: (a) using metrics of sequence stereotypy within the songs of a single bird or group to analyze sequence variability, or (b) directly comparing the song syntax of two birds or groups to obtain some measure of syntax similarity. At the core of both methods is the observation of syllable transitions and transition probabilities.

Metrics used in the first approach are sequence linearity and sequence consistency (Scharff and Nottebohm 1991). Sequence linearity quantifies the possible transitions that can be observed after each unique syllable of song and is calculated by dividing the number of unique syllables by the number of syllable transitions. In a completely linear song, each syllable has only one transition. Sequence consistency quantifies the frequency with which the dominant syntax occurs (i.e., how often the most common path through possible syllables is actually taken) and is calculated by summing the dominant transition probabilities for each syllable and dividing by the sum of all transition probabilities. Sequence stereotypy is defined as the average of the linearity and consistency scores. Newer versions of sequence stereotypy measure base scores on the entropy of the transition probability distribution for each syllable (Haesler et al. 2007; Sakata et al. 2008).

The direct comparison of the syntax of two birds/groups is a trickier proposition than comparisons of syntax variability. This approach must account for phonetic content of the song, assess sequence variability, and find a metric to combine these two measures. Funabiki and Konishi (2003) generated syntax similarity scores by first calculating the transition probability for every pair of syllables and then assessing the amount of overlap between the probability distributions for each syllable in the tutor and pupil song. These overlaps are summed and weighted by syllable frequency in the pupil's song. These scores are only valid if the experimenter is sure that syllable "X" in the tutor song is the same as syllable "X" in the pupil song. The authors addressed this by using a version of software called Sound Analysis Pro (SAP; Tchernichovski et al. 2000), developed for analysis of zebra finch songs. SAP calculates similarity scores based on spectral features of the syllables, including frequency modulation, pitch, pitch goodness, and entropy. Syllables in the pupil's song were scored for phonetic similarity against syllables in the tutor's song, and the above procedure performed on matched syllables. The latest version of SAP

includes a function for measuring sequential similarity as well. We are implementing a variation of Funabiki and Konishi's procedure in our own studies. We start by generating transition probability distributions for each syllable, resulting in a matrix of probabilities for each bird/group. Each matrix row represents the distribution for a single syllable. The correlations between corresponding syllable distributions are calculated and weighted by the phonetical similarity score (from SAP) for the relevant syllable. These correlations are summed and divided by the number of syllables in the pupil's song to produce a similarity score. Syllables present in the pupil's song but not the tutor's, or vice versa, yield correlations of 0 and thus bring the score down. Again, all of these similarity measures depend on reliable phonetical scoring and may thus become problematic when studying phonetically disrupted song.

Distinct from quantification of variability or similarity, one group uses computational models to analyze the complex syntax of Bengalese finches (Kakishita et al. 2007). Multiple song units (motifs) are contained within a bout of Bengalese finch song. Each motif can be separated into small chunks of stereotyped sequences. Thus, Bengalese finch song can be said to possess a hierarchical structure. The authors use the term "double articulation" by analogy with language, but this term implies the organization of meaningless units (phonemes) into meaningful units (morphemes), which are then organized into meaningful strings (sentences). Based on the arguable assumption that Bengalese finch song is *k*-reversible (see Angluin 1982), the authors used an automata induction approach to model song syntax, with the final model represented as an N-gram of chunks. This seems an effective method for reducing complex syntax down to a relatively simple model, perhaps paving the way for novel analysis methods.

Modeling

A physiologically based associative learning model of the song circuit (Troyer and Doupe 2000) speaks to questions of network activity addressed above. Here we focus on findings relevant to syntax learning. Based on the model's behavior, song sequence generation results from a reciprocal sensorimotor interaction between the two populations of HVC projection neurons, with the motor component encoded in HVC_{RA} neurons, and the sensory component encoded in HVC_{X} neurons. Further, the AFP modulates activity in RA via a reinforcement signal, with RA activity biased by AFP input to more closely match syllable transitions in the tutor song. The AFP teaching signal is calculated based on information about the current motor output, received from HVC. The ability of the AFP to influence RA activity is predicted to be maximal during the peak period of sequence learning, with HVC_{RA} projections dominating RA input after learning has occurred.

Fiete and colleagues (2004) constructed a simple three layer feedforward model of HVC→RA→vocal output to investigate the role of HVC sparse

coding on sequence learning. Essentially, the network produces some output that must be matched to some desired output through adjustments of the weights on connections between HVC and RA. The overall learning speed of the network decreased with increasing numbers of HVC bursts per motif under both gradient descent and reinforcement learning algorithms, due to increasing interference in the weight updates for different synapses. Thus, if HVC activity is sparse, synaptic interference is reduced. This interference is minimized if each HVC→RA synapse is used only once per motif, exactly what was observed biologically in awake behaving birds by Hahnloser et al. (2002).

A two-compartment physiological model of HVC_{RA} projection neurons has been used to investigate the role of intrinsic bursting in the generation of the sparse firing sequences observed in these cells (Jin et al. 2007). Simulations were first run without intrinsic bursting. While burst sequence generation occurred under these conditions, the network was highly sensitive to initial conditions of the simulation and could become unstable due to runaway excitation. After introducing intrinsic bursting properties into the HVC_{RA} neurons, runaway instability was abolished and the network became more robust. Intrinsic bursting in the model is driven by dendritic calcium spikes, a phenomenon yet to be observed in real HVC_{RA} neurons.

Other Species

This chapter has necessarily focused on songbirds for probing the neurobiological basis of vocal learning because so few other animal groups have the capacity for mimicry or the creation of new sounds via the vocal apparatus. Of those that do, fewer still are amenable to physiological experiments. In addition to songbirds, parrots and hummingbirds, certain species of bats, marine mammals such as whales, and elephants have been documented to possess this ability (Bougham 1998; Janik et al. 2006; Suzuki et al. 2006; Poole et al. 2005). Importantly, among primates, only humans are vocal learners. Changes in monkey or ape call structure are largely attributable to maturational factors, with a primary predictor of these changes being weight (Hammerschmidt and Fischer 2008). Some minor acoustic modifications not associated with weight have been documented. In the laboratory setting, cotton-topped tamarins exposed to loud white noise can make non-syntactical changes in the duration, timing, and amplitude of their calls to avoid the interference (Egnor et al. 2007). In natural settings, subtle modifications appear to enhance the match between vocalizations within given populations of Barbary macaques (*Macaca sylvanus*) (Fischer 2003). This latter phenomenon has been likened to speech accommodation in humans, where even subtle changes in syntax are observed as a person subconsciously adjusts their speech pattern to match that of his or her conversant (Giles 1984). How auditory feedback is used to accomplish vocal accommodation remains an open question. Again, these changes are on a much smaller scale than those observed in "bonafide" vocal learners.

Though limited in their ability to create new vocalizations, nonhuman primates can learn to associate certain calls with meanings, as can virtually all animals. The calls of vervet monkeys (*Chlorocebus aethiops*) appear innate: young vervets sound much like mature ones. However, a young vervet requires experience to use its vocalizations in the appropriate context and to interpret conspecific calls correctly (Seyfarth et al. 1980). A question of syntax occurs when combinations of calls convey new meaning. Putty-nosed monkeys (*Cercopithecus nictitans*) make two different alarm calls, termed "hacks" and "pyows," in response to distinct predators. A series of pyows appears to signify leopards, while hacks or hacks followed by pyows are a response to eagles. At other times, males produce 1–4 pyows followed by 1–4 hacks. These series reliably occur prior to movement of the group (Arnold and Zuberbühler 2008). The joining of two distinct alarm calls to convey a new meaning would appear to fit the bill for syntactic recombination. However, to be truly compositional, the meanings of the individual calls must be relevant to their combinatorial meaning (Hurford 2009), a finer but critical point that remains to be tested.

The possibility of compositionality arises from studies by Zuberbühler (2002) of Diana monkeys (*C. diana*) who listen in on sympatric Campbell's monkey calls (*C. campbelli*). Like vervet monkeys, Campbell's monkeys make alarm calls to nearby predators, and these are distinct for leopards versus eagles. When Diana monkeys hear the Campbell's calls, they make their own corresponding alarm calls. Campbell's males make a different sound when predators are at a distance, or when some other less critical disturbance occurs. In these situations, they emit a pair of low-pitched "boom" vocalizations. Strikingly, when these booms precede a Campbell's alarm call, Diana monkeys no longer make their own alarm calls. The addition of a boom appears to serves as a new, cross-species semantic signal, along the lines of "pay no attention to the man behind the curtain," or, as Zuberbhühler notes, something akin to human linguistic hedges such as "kind of" or "maybe." Though not proven, the latter interpretation raises the possibility of compositionality since the boom can be interpreted as adding a "not to worry" signal to the "there's a predator" one.

At the extreme of nonhuman primate call complexity is gibbon song. *Hylobates agilis* produce songs that are organized into complex sequences of several call phases. Females' calls differ acoustically between individuals, opening the potential for use in individual recognition (Oyakawa et al. 2007). At least in *H. lar* and *H. pileatus* however, such calls are inherited rather than learned. Hybrid offspring produce hybrid vocalizations in that the acoustic features are intermediate between those of each parent (Geissman 2000). Reminiscent of the putty-nosed monkey calls described above, certain components of gibbons' songs are differentially emphasized in different contexts. Specifically, white-handed gibbon songs were shorter, contained fewer sharp "wow" notes and shorter "hoo" notes when in a predatory context than when duetting. Differences in these song patterns appeared salient to the few members that temporarily left the group during the observation period. Upon returning, males responded

with their own songs after hearing the groups' songs to predators, but not after hearing duets (Clarke et al. 2006).

As summarized by Hammerschmidt and Fischer (2009), "While vocal production appears largely innate, learning does play a role in the usage and comprehension of calls, but this is not restricted to the primate order." Further, recombination of largely innate calls can be used to convey new meaning, but it remains unclear that the new "whole" is a manifestation of the sum of its parts, or indeed, has any shared meaning with them.

Returning to vocal learners, in a few cases, individual elephants (*Loxodonta africana*) have been observed to mimic nonspecies specific sounds (including human words!) (Poole et al. 2005). While the uniqueness of an elephant's trunk has been compared to singularity of human language (Pinker 1994), the discovery that elephants also possess the rare ability of vocal learning is quite recent. Investigation into the ontogeny of elephant's species-specific signals is even more nascent (Stoeger-Horwath et al. 2007). It will be interesting to learn whether elephants combine and respond to mixes of seismic and vocal communication signals in any sort of compositional manner (O'Connell-Rodwell 2007). Among marine mammals, hump-backed whales (*Megaptera novaeangliae*) produce some of the most elaborate songs in the animal kingdom. An individual whale sings its own distinctive song, with other whales in the same population singing similar ones. Songs recorded across 30 years reveal changes that are rapidly adopted within a given population. Changes occur mainly in the middle of the breeding season and are correlated with increased durations of the song sessions, suggesting that the song changes are part of a sexual display (Payne 2000).

Analysis of song structure in whales reveals a series of units referred to as a phrase and an unbroken sequence of similar phrases, called a theme (Payne and McVay 1971). Within a song there are variations on a phrase; slight changes occur across successive renditions. Themes are sung in an ordered sequence and compose a song which can last half an hour. Songs are strung together in sessions which, themselves, can last for hours. The use of information theory techniques (Suzuki et al. 2006) has confirmed the earlier proposal (Payne and McVay 1971) that the songs exhibit hierarchical structure. However, the claim that humpback whales can sing an infinite number of songs from a finite set of units, startlingly similar to human syntax, apparently overstates the case. As Hurford notes, Suzuki and colleagues combined the songs of several whales in their analyses. Simply put, a single humpback at any one time in its life only sings one song (Hurford 2009).

Finally, while marine mammals present obvious challenges for neurobiological study, neural activity in the brains of smaller mammalian vocal learners has been examined. Mustached bats (*Pteronotus parnellii*) can combine otherwise independently emitted syllables. Esser and colleagues (1997) used playback of these naturally occurring heterosyllabic composites as well as temporally destructured versions to determine the response properties of neurons. They

uncovered regions in nonprimary auditory cortex that were sensitive specifically to the composite structure of communication calls. It has been proposed that the limited number of vocal gestures available to nonhuman primates has driven the ability of receivers to process signal combinations (see Számadó et al., this volume). Detection of combination-sensitive neurons in auditory association cortex in bats, also detected in songbird pallium (Margoliash and Fortune 1992), may provide the neural building blocks for more complex syntactical processing in humans.

Mirror Neurons

Any communication system requires individuals to attend to and understand the actions of others, be they gestural or vocal. A possible neural substrate for such has been well studied in primates and has recently been identified in songbirds. Mirror neurons were originally observed via neurophysiological recordings in monkey premotor cortex (Rizzolatti and Arbib 1998). These neurons fired not only when subjects grasped or manipulated objects, but also when the monkey observed the experimenter making a similar gesture. Some mirror neurons are highly specific, not coding only for an action, but also how that action is executed. For example, they fire during observation of grasping movements, but only when the object is grasped with the index finger and thumb. They can fire when the grasping action is performed with the mouth as well as with the hand (Gentilucci and Corballis 2006). In general, mirror neurons are proposed to code for representations of actions, which can then be used for imitating and understanding the actions of conspecifics. Evolutionarily speaking, they may have been instrumental in the transfer of the gestural communication system from the hand to the mouth (Gentilucci et al. 2001). Based on imaging studies of the KE family, it has also been speculated that *FOXP2* might play a role in the incorporation of vocal articulation into the mirror system (Gentilucci and Corballis 2006).

Along with substantial evidence for mirror neuron system (MNS) dysfunction in autism (Iacoboni and Dapretto 2006), the human MNS has also been hypothesized to play a role in apraxia, a cognitive motor disorder in which the patient loses the ability to perform learned, skilled actions accurately. The most common form, ideomotor apraxia, has been described as "an impairment in the timing, sequencing, and spatial organization of gestural movements." [Source of quote?] Patients with ideomotor apraxia cannot tell if someone else is performing an action correctly or not (McGeoch et al. 2007). Functional MRI has been used to investigate the role of the human MNS in representing hierarchical complexity during the observation of action sequences (Molnar-Szakacs et al. 2006). Observation of object manipulation sequences recruited classic MNS regions, and MNS activity appeared to be modulated by the perceived motor complexity of the action. These results support Arbib's theory of

language evolution (Arbib 2005), wherein language is thought to have evolved out of the motor system for gestures, and provide a connection between developmental and neural evidence linking motor and language functions.

Neurons displaying precise auditory–vocal correspondence (i.e., mirror neurons) have recently been observed in the song system of swamp sparrows (*Melospiza georgiana*) (Prather et al. 2008), the first identification of auditory–vocal correspondence in single neurons. HVC_X neurons displayed highly selective auditory responses, typically activated by only one song type. These responses were sparse, occurring at a precise phase in a given syllable. The same neurons also fired selectively when the bird sang the preferred song type, also phase locked to a particular part of a given syllable. The singing-related activity of these neurons was motor related and not due to auditory feedback of the song. Interestingly, the auditory responses of HVC_X neurons extend to the songs of conspecifics with note sequences similar to that of the bird's own preferred song type.

Summary

For those interested in the biological origins and evolution of language, a daunting obstacle is the lack of neurobiological data on users of protolanguage and early language. Making matters more challenging, humans are the only current language users on the planet. Thus a comparative approach is necessary. While no complete animal model of human language exists, animal models can be used to investigate the neural systems underlying different aspects of language. This includes investigation of molecules identified in genetic studies of human language disorders. Animal models can also be useful in studying the biological basis of linguistic preadaptations, such as the motor and perceptual skills needed to learn sequential ordering of vocal utterances.

Songbirds provide an ideal system for such studies: they learn vocalizations under similar constraints to humans learning language, possess identified neural circuits underlying this ability that are similar to circuits in humans important for language, and share some genetic basis for vocal learning with humans (Jarvis 2004). From a practical standpoint, their behavior is easily quantifiable, and they are amenable to laboratory life. Admittedly, birdsong is far from being compositionally semantic. To stress the absence of meaning upon combination of its units, we have described the simple structure of their songs as exhibiting only phonetical syntax. However, the experience-dependent shaping of neural circuitry for song may point to building blocks for more complex micro- and macro-circuits that comprise human language centers. Perhaps just as similar selection pressures drove parallel evolution of the eye in dozens of distinct lineages (Land and Fernald 1992), a similar situation may likely hold for parallel biological solutions to the problem of learned vocal communication.

To date, FOXP2 is the only single molecule to be repeatedly linked to language (Marcus and Fisher 2003). Here, we argue that what we learn about FOXP2, from humans as well as animal models, can be leveraged as a molecular wedge into the networks underlying language. Studies in songbirds (Haesler et al. 2007) and the more genetically tractable mice (Enard 2002; Teramitsu and White 2008) suggest that FoxP2 is important for species-typical procedurally learned behaviors, such as locomotor skills in rodents and song in birds. These reports also highlight the role of the striatum in the experience-dependent neural changes underlying such skills. We predict that a comparative network analysis of the gene targets of FOXP2, identified via deep sequencing or microarray analyses, will reveal some shared connectivity related to basic skill learning, as well as connections unique to the species' skill sets including speech, and possibly language, in humans. In this light, it will be interesting to see whether CNTNAP2 is a FoxP2 target in species other than humans (Vernes et al. 2008).

Learned vocal motor control gets us only part of the way to human linguistic syntax. The additional capacity of conveying and processing semantic content is key to moving beyond the musical-like realm of birdsong to semantically compositional language. Discoveries on the biological basis and evolution of both vocal and nonvocal learners' capacity for symbolic representation and theory of mind must be joined with the findings reviewed here. In this vein, mirror neurons provide a hypothetical link between an action and its meaning, or intent, when committed by one's self or by others. Whether or not the mirror neuron system is part of this particular puzzle, somewhere along the hominid lineage, neural systems for complex meaning must have intersected with those for ordering of vocal output to lay the basis for human linguistic syntax.