



Targets for a comparative neurobiology of language

Justin T. Kiggins¹, Jordan A. Comins² and Timothy Q. Gentner^{1,2,3*}

¹ Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA, USA

² Department of Psychology, University of California, San Diego, La Jolla, CA, USA

³ Kavli Institute for Brain and Mind, University of California, San Diego, La Jolla, CA, USA

Edited by:

Angela Dorkas Friederici, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:

Paul M. Nealen, Indiana University of Pennsylvania, USA

Jonas Obleser, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

*Correspondence:

Timothy Q. Gentner, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, MC0109, La Jolla, CA 92093
858-822-6763, USA.
e-mail: tgentner@ucsd.edu

One longstanding impediment to progress in understanding the neural basis of language is the development of model systems that retain language-relevant cognitive behaviors yet permit invasive cellular neuroscience methods. Recent experiments in songbirds suggest that this group may be developed into a powerful animal model, particularly for components of grammatical processing. It remains unknown, however, what a neuroscience of language perception may look like when instantiated at the cellular or network level. Here we deconstruct language perception into a minimal set of cognitive processes necessary to support grammatical processing. We then review the current state of our understanding about the neural mechanisms of these requisite cognitive processes in songbirds. We note where current knowledge is lacking, and suggest how these mechanisms may ultimately combine to support an emergent mechanism capable of processing grammatical structures of differing complexity.

Keywords: comparative neurobiology, language, songbirds

INTRODUCTION

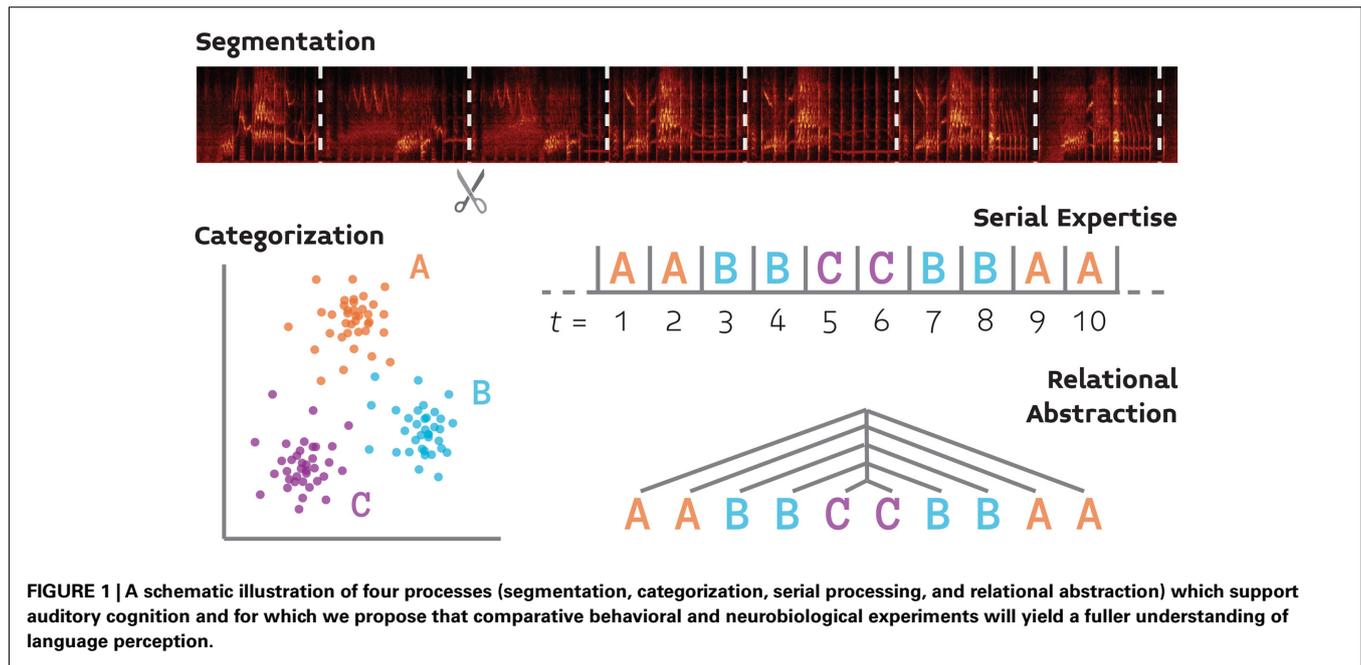
Communication is ubiquitous among animals, but only humans seem to possess language. The uniqueness of modern language among animal communication systems has fostered broad, often contentious, inquiries regarding its evolutionary origins through either adaptation or exaptation, along with attempts to define a subset of unique, language-specific cognitive abilities. Work along these lines has focused largely on whether language exists along some continuum with other communication systems, or is categorically distinct (Hauser et al., 2002; Fitch and Hauser, 2004; Fitch et al., 2005; Jackendoff and Pinker, 2005; Pinker and Jackendoff, 2005; Margoliash and Nusbaum, 2009; Berwick et al., 2011; Terrace, 2011), and attempted to dichotomize cognitive processes into those that are or are not “human-like” (Jackendoff and Pinker, 2005; Pinker and Jackendoff, 2005). While recognizing the importance of this work, we propose a different approach to understanding the current neural mechanisms and evolution of language. Rather than identifying putatively unique, language-relevant abilities and asking whether non-human animals show evidence for them, we outline a set of cognitive abilities that are unquestionably shared by many animals but which are nonetheless prerequisite to human language. We seek to guide both the language evolution and neurobiology conversations toward more fundamental auditory and memory challenges that many vocal communication systems share. We suggest that studying these more basic processes will yield in the near term to a biological understanding of these processes with neuronal and network-level resolution. Such knowledge will constitute an initial substrate for an ultimately more complete neurobiology of language, provide a clearer picture of the mechanisms available in proto-languages and/or ancestral hominins, and a biological context within which models of putatively unique language mechanisms

can be generated and tested. In short, we propose that there is more to be learned about the neurobiology and evolution of language by studying mechanisms that are shared, rather than those that are unique.

We review four basic components of auditory cognition (Figure 1) that follow the foregoing reasoning, and for which the basic behavioral and neurobiological groundwork has already been laid. This list, which includes segmentation, serial expertise, categorization, and relational abstraction, is not meant to be exhaustive, but rather demonstrative of the proposed approach. We focus our discussion of the neurobiology of these processes on songbirds because this system provides the most well-developed model for the neurobiology of vocal communication, and thus will have much to contribute (at least initially) to a comparative neurobiology of language.

SEGMENTATION

A fundamental aspect of perceiving complex communication signals such as speech and language is the ability to segment a sound into temporally distinct auditory objects. The variations in acoustic pressure that are encoded by the cochlea are continuous in time, whereas the useful units of language are discrete segments of this continuous stream. Segmentation occurs at various hierarchical timescales. For speech, this includes phonemic, syllabic, and morphemic boundaries, while for birdsong this includes note, syllable, and motif boundaries. Human infants are remarkably good at detecting word boundaries in speech. At only 8 months of age they can detect word boundaries from fluent speech following a mere 2-min of exposure to word-streams (Saffran et al., 1996). Evidence suggests that segmentation occurs largely through transition statistics and prosodic cues (Jusczyk, 1999).



SEGMENTATION CUES

One reliable cue for segmenting sounds into discrete elements is through an evaluation of the joint statistics of neighboring acoustic elements. Frequently co-occurring pairs of elements are unlikely to be a boundary between segments, while improbable pairs of elements likely represent a boundary. These sequential probabilities can be learned by adult and infant humans very quickly. When exposed to continuous sequences of “nonsense” words, constructed such that the joint probabilities of syllables are high within words and low at the boundaries, listeners can identify novel words that conform to the sequential probabilities learned previously (Saffran et al., 1996). The neural mechanisms of statistical segmentation are, however, likely domain general, as statistical segmentation is not exclusive to speech segmentation, but is also involved in the segmentation of tone sequences (Saffran et al., 1999).

In addition to transition statistics, there are other perceptual cues that human adults use to detect word boundaries. In adulthood, these prosodic cues are relied on very heavily for humans to segment speech. Humans begin showing sensitivity to prosodic cues, such as stress regularities, during infancy (Thiessen and Saffran, 2003). Indeed, event-related potentials in infants as young as 4-months show mismatch responses for disyllabic words with alternative stress patterns from those of their parents’ native language (Friederici et al., 2007). Recent investigations show adults can isolate words from fluent non-native speech using solely prosodic cues (Endress and Hauser, 2010), providing evidence for some aspects of prosody being universally accessible in segmentation.

Further evidence from research on music perception that word segmentation interacts with auditory perceptual grouping, again suggesting general cognitive processes for segmentation. Listeners group tone sequences consistent with the grouping of syllable durations in the listener’s native language (Iversen et al., 2008).

This influence of native language prosody on perception in an auditory but decidedly non-linguistic domain further supports the potential that more general acoustical segmentation processes are being utilized in speech segmentation.

NEUROBIOLOGY OF SEGMENTATION

Much of the human research on the neuroscience of segmentation has relied on functional imaging to localize anatomical areas involved in statistical segmentation and prosodic processing and on extracranial electrophysiology (EEG) to elucidate the temporal dynamics of processing. The N100 and N400 components of the event-related potential have been implicated, respectively, in word segmentation (Sanders et al., 2002) and statistical learning (Abla et al., 2008). Functional imaging work has implicated the superior temporal gyrus (STG) in analysis of transitional probabilities (McNealy et al., 2010), with left STG activation also correlated with individual’s performance on the discrimination of trained pseudo-words. Further, posterior left inferior and middle frontal gyri (IFG/MFG) showed stronger activation for “words” (highly probable trisyllabic strings) versus strings with lower probability. In both cases, STG and I/MFG activation was reinforced by prosodic cues. These types of studies do provide information about neural sensitivity and coarse anatomical compartmentalization, but no information about the neuronal circuit mechanism underlying the perception of prosodic cues and sequence statistics or how these are integrated to determine segmentation boundaries.

PERCEPTUAL SEGMENTATION OF BIRDSONG

Like human speech, birdsong is composed of hierarchically organized sound elements, though the specific elements and their hierarchical organization vary across species. Zebra finches sing a rather stereotyped song consisting of temporally distinct elements called “notes,” typically lasting tens of milliseconds in duration. Like zebra finches, Bengalese finches utilize temporally distinct

notes in their song, but with more probabilistic transition statistics, resulting in more variable song. Starling song is even more complex, both in terms of spectrotemporal acoustic features and the sequencing of song elements. The diversity in song across songbirds has allowed researchers to explore various aspects of segmentation behaviorally. In all songbirds, their respective elements have boundaries and associated statistics and, like humans, songbirds show sensitivity to both statistical and prosodic cues in acoustical segmentation.

Segmentation in language perception occurs on multiple timescales, as phonemic, syllabic, and morphemic boundaries must be defined. European starlings, one species of songbird, show perceptual sensitivity to manipulations of conspecific song on multiple timescales. One of the major elements of starling song is the motif, a spectrotemporally complex yet stereotyped vocalization. A single male starling will have dozens of unique motifs in its repertoire, which will be combined into long bouts, which last on the order of a minute. The combination of stereotyped elements and more variable strings of elements yield unique statistics on both the intra and inter-motif timescales which are available for recognition (Gentner and Hulse, 1998; Gentner, 2008). After being trained to classify song excerpts from different singers, starlings show sensitivity to both intra-motif temporal statistics and inter-motif statistics. When sub-motif acoustical features are permuted, thus breaking down intra-motif statistical relationships, birds show a deficit in recognition. When asked to classify excerpts from the same birds but composed of novel motifs, however, birds perform equally well for excerpts with intact motifs and intra-motif permutations (Gentner, 2008). In all, these results indicate that starlings are sensitive to the statistical relationships within motifs. Starlings are also sensitive to the ordering of motifs in a similar recognition task. After training starlings on a vocal recognition task, new song excerpts were generated which either maintained only the frequency of occurrence of motifs or where both frequency and first-order transition statistics were maintained. Though still above chance, starlings were far more impaired at recognizing songs lacking first-order transition statistics (Gentner and Hulse, 1998). Though this work demonstrates that starlings are sensitive to learned statistics of other starlings' song, it is not clear whether they use these statistical cues or other acoustical features in grouping and segmentation.

One way that segmentation has been explicitly assessed in songbirds is by investigating the way that tutor song elements are incorporated into juvenile songs during sensorimotor song learning. Young birds learn songs composed of smaller units and importantly, like human infants, show evidence of segmenting via the use of some prosodic cues. Prosodic cues for defining segmentation boundaries were explored by exposing young zebra finches to multiple tutors and analyzing how tutor song elements were grouped in juvenile song. During learning, the juveniles incorporated segments of up to eight elements from the multiple tutors, integrating the tutors' song into their own by grouping elements together that were grouped in the tutor song. These segmentation boundaries were more likely to occur where the tutor produced longer inter-element silent periods (Williams and Staples, 1992). While silent periods are not a reliable cue for speech segmentation (as noted earlier, stress regularities and other cues are more

important), the implication of at least one acoustic feature in segmentation of learned motor sequences suggests that there may be other untested acoustic cues that carry prosodic information in songbirds and these specific cues might vary across species. Indeed, as described above, starlings are sensitive to motif sequencing, but the boundaries between motifs often do not contain silence. Further research will be necessary to explore the reliance on various prosodic cues for segmentation.

As zebra finch song is relatively static in its sequencing, the more variable transition statistics of Bengalese finch song offers an opportunity to explore juveniles' reliance on statistical transition cues in addition to prosodic cues. Like zebra finches, when juvenile Bengalese finches are exposed to multiple tutors, longer inter-element silent periods predicted segment boundaries. Further, Bengalese finches more readily group tutor song elements with higher transitional probabilities and segmentation boundaries in juvenile song are associated with lower transitional probabilities in the tutor songs (Takahasi et al., 2010). In both of these experiments, however, the cues that juveniles can rely on are largely uncontrolled by the experimenters and are limited by the cues that are being produced by the tutor birds. Additional evidence for the role of pairwise statistical relationships in song learning comes from work in white crowned sparrows. By comparing young birds exposed to only pairs of elements from the adult tutor song or elements in isolation, researchers showed that, though elements were grouped in both sets of birds, those birds which had access to paired elements during learning grouped such pairs together in their mature song (Rose et al., 2004). It is interesting to note, however, that white crowned sparrows grouped elements together even when they only learned those elements in isolation. Nonetheless, the transition statistics between pairs of elements was a strong cue for learned grouping. In order to better understand the relative contributions of prosodic and statistical cues in the segmentation of strings in sensorimotor learning, further studies will be necessary training young songbirds on synthetic song, carefully manipulating the availability of prosodic (i.e., inter-element silent periods, syllable durations, syllable stress) and statistical cues independently (Lipkind and Tchernichovski, 2011). Such an approach would also enable exploring a battery of other prosodic cues that have been implicated in human segmentation.

In addition to segmentation in sensorimotor learning, which must be supported by both perceptual and motor processes and thus could be constrained by either, further research is necessary to explore perceptual contributions to segmentation independently from grouped motor behaviors. Utilizing standard operant conditioning and habituation tasks, which songbirds are amenable to, an important direction for future research will be to further explore the perceptual sensitivity to transition statistics and prosodic cues for segmentation and the relative weight placed on the two types of cues when birds must parse strings in order to perform a behavioral task.

NEUROBIOLOGY OF BIRDSONG SEGMENTATION

Given that songbirds show sensitivity to both prosodic cues and sequence statistics in vocal learning, they offer a unique opportunity to understand the neural mechanisms of these processes, and yet, the neurobiological mechanisms of segmentation have

not been studied extensively in songbirds. Is there evidence of segmentation in the avian forebrain? Though the neurobiology of segmentation has not been explored directly, there is extensive evidence for neural processing of acoustic features typically associated with prosody and some evidence for experience-dependent neural representations of the experienced statistics of element transitions.

As we ascend the avian auditory forebrain, we find that traditional spectrotemporal receptive field models from spike-triggered averaging are increasingly poor at predicting neural responses and neurons become much more sensitive to the acoustical features of conspecific vocalizations (Theunissen et al., 2000; Gill et al., 2008). For example, in CMM of the starling, it has been demonstrated that units respond to extracted acoustical features of starling song and that responses to full vocalizations can be predicted by the responses to isolated features (Meliza et al., 2010). As noted above, the only acoustic feature which has been implicated as a prosodic cue for songbirds is silence. This makes songbirds a good model to evaluate theories of neural encoding and perception of time.

Though neural representations of prosodic acoustical features is somewhat trivial in terms of feed-forward hierarchical processing of sensory features, the processing of sequence statistics poses a more challenging theoretical problem. One clue may lie in temporal combination selectivity, the tendency for some neurons to respond very selectively to specific pairings of song elements. This has been demonstrated in Field L and HVC of the zebra finch (Margoliash, 1983; Margoliash and Fortune, 1992; Lewicki and Arthur, 1996) and more recently in the Bengalese finch (Nishikawa et al., 2008). As the neural representation of pairs of elements is particularly relevant to our understanding of the neural mechanisms of serial expertise, we will discuss temporal combination selectivity at length later. Given these results, it would be worthwhile to explore the extent to which temporal combination sensitivity supports learned segmentation of song. To what extent do transitional statistics and prosodic cues influence temporal combination sensitivity in auditory and premotor areas? In addition to past work in primary auditory (Field L) and premotor (HVC) structures, to what extent do secondary auditory areas such as NCM and CM show sensitivity to temporal combinations of song elements? And how do prosodic and statistical cues differentially contribute to temporal combination sensitivity through this hierarchy of auditory processing?

SEGMENTATION: CONCLUSIONS

There are many perceptual cues humans and songbirds can use to detect element boundaries and segment communication signals. Evidence for some aspects of speech segmentation in humans being domain-general processes and birds utilizing similar strategies in segmenting human and avian communication signals indicates that birds may be useful experimental subjects in eliciting the underlying neural computations and biological mechanisms of these processes.

SERIAL EXPERTISE

REPRESENTATIONS OF SERIAL ORDER IN HUMANS

Sensitivity to the ordering of linguistic units across time is vital to language comprehension. Indeed, in many languages, word order plays a vital component in assessing grammaticality (e.g., English)

while in others syllable-order serves an important role in phonology as in determining stress (e.g., Polish). As such, a requisite capacity for language is knowledge of the serial order of events occurring at multiple timescales within a signal. Thus, a fundamental challenge to linguistic processing is monitoring not only which elements occur in a sequence, but also where they occur. A system capable of linguistic processing must have at its disposal sufficient memory to store multiple items after a signal fades and to represent the serial arrangement of those items. Understanding how temporal pattern information is encoded, notoriously christened the “problem of serial order” by Lashley (1951), has been of longstanding importance to psychology since Ebbinghaus’ early models of the serial position effect (i.e., primacy and recency). Here, we briefly discuss the two most prominent behavioral accounts of sequence-encoding: chaining and positional models.

Chaining models emerged from the classic stimulus–response theories of serial behavior championed by Watson (1920), Washburn (1916), and Skinner (1934). These models propose that a given element’s location in a sequence is encoded by association with both the preceding and succeeding element. Accordingly, the sequence ABCD would be encoded (most simply) as a sequence of pairwise associations, such as A–B, B–C, C–D, where the recall of a single item initiates the recall of a subsequent item. Positional models, on the other hand, suggest that items are encoded on the basis of their position in the sequence. The limitations in understanding serial behavior and learning using only strict associative chaining theories, even in regards to language, have long been known (Lashley, 1951) and a considerable amount of research has been undertaken to demonstrate that sequence learning in humans and non-humans additionally relies on positional information. Further, while both models posit potential psychological accounts to encode serial order, positional models do a better job accounting for common errors in human serial recall. For example, intrusion errors occur when trying to recall one of two lists, such as ABCD and EFGH. Such errors occur when an item from the second list is mistakenly presented during recall of the first list, but in its canonical position assignment (i.e., AFCD). Chaining models would predict that an error in the second position would trigger a cascade of incorrect responses, such as AFGH, a type of mistake not commonly seen during erroneous sequence recall (Henson et al., 1996). Given that human errors during serial recall suggest an encoding method that can incorporate positional cues, one might ask how these positional representations are instantiated. The most compelling behavioral account suggests that positions are assigned relative to certain perceptual anchors. Specifically, Henson’s start-end model states that positional assignments are encoded relative to sequence edges (Henson et al., 1996).

Further evidence supporting an edge-based serial processing system in human adults comes from the study of artificial grammar learning (AGL). For instance, Endress et al. (2005, 2009a,b) demonstrated that repetition-based rule structures are only generalized if those repetitions occur on the edges of a sequence. For example, syllable sequences of the form ABCDDEF were not generalized by subjects. If, however, the repetitions were edge-based, such as ABCDEFF, subjects did effectively generalize knowledge of the repetition rule. Importantly, subjects could still discriminate

grammatical and ungrammatical sequences with either internal or edge-based repetitions, suggesting differences in generalization ability were not based on a failure to detect internal repetitions. Thus, the authors conclude that subjects relied on both detecting a repetition and processing where the repetition occurred, thereby paralleling the claim of Henson that items in initial and final positions are more robustly encoded than those at inner positions of a sequence. Such reliable encoding of sequence edges thus appears to exploit more primitive memory processes that enhance the representation of early and late items in a chain, a hypothesis supported by AGL work in non-human primates (see Serial Expertise in Non-Human Animals; Endress et al., 2009a, 2010)

The use of positional information to recall serial order is not purely an artifact of sequence memory or AGL tasks (Endress et al., 2005). Rather, natural languages possess many grammatical and phonological properties that rely on the positional assignment of elements within a sequence. For example, affixation rules most often alter the beginning or end positions of a sequence, rather than middle positions (Endress et al., 2009b). Additionally, prosodic components of language similarly rely on positional information, as in allocating syllabic stress (e.g., in Hungarian, the first syllable of a word is stressed; in Macedonian, the antepenultimate; in Polish, the penultimate; in French the ultimate). Thus, the convergence of positional encoding strategies for serial recall, as well as their prevalence in natural languages, suggests that such serial expertise relying on positional, as well as chaining, strategies is a necessary cognitive ability used for linguistic communication.

SERIAL EXPERTISE IN NON-HUMAN ANIMALS

To have an appropriate animal model for grammatical processing, a species must demonstrate similar working memory constraints as humans for sequence recall as well as similar encoding strategies. While several studies of serial behaviors in pigeons and mammals are easily explained exclusively by chaining strategies (Weisman et al., 1980; Balleine et al., 1995), others are not (Straub and Terrace, 1982; D'Amato and Colombo, 1988; Endress et al., 2010). Some of these latter experiments demonstrate reliance on positional information during serial recall via the "simultaneous chaining" procedure (Terrace, 2005). In this task, animals learn to touch images in a specific order, receiving reinforcement only if the entire sequence is recalled correctly. Crucially, the images are presented simultaneously but in a random location for each trial. Consequently, unlike successive chain tasks, where serial learning was assessed with a specified spatial configuration (i.e., in a maze), the simultaneous chain paradigm forces subjects to acquire a representation of stimulus serial order independent of learning a fixed set of motor responses. This procedure has provided strong support for the use of positional information to learn and recall serially ordered visual stimuli (D'Amato and Colombo, 1988; Chen et al., 1997). In addition, a more recent experiment undertaken by Endress et al. (2010) found that both chimpanzees and humans similarly utilize edge-based positional information in an auditory artificial grammar habituation/dishabituation task.

Amongst vocal learning species, songbirds have been the most extensively studied regarding their serial processing capacities. Through a direct comparison with human subjects, one species of songbird, the European starling (*Sturnus vulgaris*), has

demonstrably similar auditory memory capacity limits and decay functions for short-term store (Zokoll et al., 2008). In addition, an experiment by Comins and Gentner (2010) explored the sequence-encoding strategies of this same species of songbird and reported that starlings rely, at least partially, on absolute and relative position information for representing serial order.

This body of work indicates that animals, like humans, can use a host of positional cues in the absence of associative chains to recall serial order. Unlike other non-human species, only the songbirds undergo a vocal learning procedure with many notable similarities to human infants acquiring knowledge of language (Brainard and Doupe, 2002). Thus, the songbird may extend previous animal models of serial order processing and recall into a natural behavioral context tied to vocal communication and serve as a suitable window for similar processes in linguistically endowed humans.

NEURAL MECHANISMS OF TEMPORAL ORDER IN SONGBIRDS

In this section we revisit Lashley's "problem of the serial order" from a neurobiological perspective. At its core, this task requires the nervous system to enhance or suppress responses to a stimulus based on its temporal context: did stimulus event B correctly follow event A? In many communication systems, this "temporal context" can be defined on multiple levels of a hierarchy. Consider the case of speech and language, where the ordering of phonemes is crucial to the emergence of a word, and still the ordering of words can be vital to the emergence of an expression. Thus, the neural systems responsible for this behavior must integrate contextual information across a large range of timescales of stimulus events, from milliseconds to perhaps several seconds. Here, we review findings on how serial order for hierarchically organized elements of natural communication is, at least partially, represented in the zebra finch (*Taeniopygia guttata*).

Zebra finch males sing stereotyped songs to court females, who do not sing. These songs are strings of syllables and characterized by their short length and hierarchical organization. The mature zebra finch song further has a canonical syllable progression within a song bout. Thus, zebra finch song proceeds along a sequence of changing syllables, where each syllable represents a complex auditory event. Sensitivity to such serial ordering of song elements has been investigated physiologically across different levels of the avian telencephalon, namely field L and HVC (Margoliash, 1983; Margoliash and Fortune, 1992; Lewicki and Konishi, 1995; Lewicki and Arthur, 1996). Field L, analogous to the mammalian primary auditory cortex, receives thalamic afferents from the nucleus ovoidalis (Kelley and Nottebohm, 1979), while HVC is an upstream projection target of the higher subregions of field L, L1, and L3, as well as the non-primary auditory area CLM which projects to the HVC shelf. In these studies, an anesthetized male subject is exposed to several variants of his own pre-recorded song, a stimulus known to selectively drive neuronal responses particularly in HVC (Margoliash, 1986). By presenting the subject with renditions of his own song occurring in normal, reversed, syllable reversed and sub-syllable reversed orders, researchers have successfully identified classes of neurons sensitive to the progression of syllabic or sub-syllabic features or both (Margoliash and Fortune, 1992; Lewicki and Arthur, 1996). In addition, Lewicki and Arthur (1996) showed strong convergence between

the anatomical projections from field L to HVC and the sensitivity of neurons therein to higher-order temporal contexts. Specifically, recordings from the primary thalamorecipient zones of field L, L2a, and L2b, were only sensitive to temporal differences between normal and reversed song. However, HVC projecting regions L1 and L3 showed a modest percentage of temporal context sensitive cells for sub-syllable manipulations and L3 alone responded to differences in syllable order, while HVC shows an even higher proportion of neurons tuned to these stimulus properties.

Though these studies clearly show temporal sensitivity in a percentage of field L and HVC neurons, they do not directly address how these cells gate their responses to a given stimulus event as a function of its temporal context. Extracellular responses do, however, rule out simple facilitation as a potential mechanism of syllable-order sensitivity (Lewicki and Konishi, 1995; Lewicki and Arthur, 1996). Further, given that strong extracellular responses to specific syllables in forward song were nearly entirely eradicated in distorted temporal contexts, Lewicki and Konishi (1995) employed intracellular methods in HVC neurons. Their results suggest that cells sensitive to temporal context generally exhibit weak depolarization sometimes followed by inhibitory currents during the presentation of either syllables in isolation or aberrant sequential orders of syllable-pairs. When pairs of syllables are presented in their canonical order, however, this weak depolarization and inhibitory influx is followed by a nonlinear burst of action potentials. Computational models of such responses have thus predicted that zebra finch song sequencing information is organized in a chain-like manner, where nodes on the chain are responsible for variable context-sensitivities (Drew and Abbott, 2003; but see Nishikawa et al., 2008 for an alternative model in Bengalese finches).

The responses of temporal context sensitive cells in the zebra finch auditory system are highly tuned to the local and global structure of serial order of the bird's own song. Many properties of serial order representation in songbirds, however, remain to be tested. In the auditory system, the hierarchy of context sensitivity has only been studied at the level of field L and HVC, while the contribution of other auditory areas, such as CM and NCM, which show behaviorally relevant modification of song selectivity (Gentner and Margoliash, 2003; Thompson and Gentner, 2010) remains unknown. The role of NCM is particularly important in understanding temporal context sensitivity, as this is a well-established area of experience-dependent decision-making for mate decisions based on male song features in European starlings (Gentner et al., 2001; Sockman et al., 2002), whose songs are largely characterized by their motif-structure (Eens et al., 1988).

Additionally, neurophysiological explorations have yet to dissociate chaining from positional representations of sequentially arranged stimuli. For example, consider a cell that shows sensitivity to the sequence AB. At present, it is not entirely possible to isolate whether the subject is responding to B given the information provided by the association of A to B, or by B's position in the sequence relative to A (i.e., the second motif). One possible way to parse apart these types of temporal information would be to create stimuli for a subject that combine motifs across different variations of autogenous (i.e., bird's own) songs. Thus, if a bird sings two songs, one beginning with motif sequence AB and

another beginning CD, a relative position-encoding model might be robust to order violations in the sequence CB compared to BB or BC. The reason being that motif B, though presented in a non-canonical transition from C to B is still located in the correct relative position of the sequence. Such a design would dissociate between the encoding of positional versus transitional sequence information as outlined above (see Representations of Serial Order in Humans).

Finally, though this section is primarily concerned with the role of the auditory pathway on serial expertise, non-auditory areas, such as the basal ganglia, likely provide important contributions to serial order representation that remain to be explored. While the songbird anterior forebrain has been intensively studied in relation to song sequence production (for a review, see Brainard and Doupe, 2002), it has only recently been suggested as an important region of syllable-level syntax perception. Abe and Watanabe (2011) tested syntax discrimination abilities of Bengalese finches. Using an immunocytochemical technique, the authors stained Zenk protein, an immediate early gene upregulated during exposure to conspecific song (Mello and Ribeiro, 1998), to localize areas responding strongest to violations of a familiarized temporal syllable order. With this method, Abe and Watanabe (2011) found that neurons in the lateral magnocellular nucleus of the anterior nidopallium (LMAN) showed heightened activation to temporal orders. LMAN, along with two other regions in the anterior forebrain pathway, Area X (which receives projection from HVC) and the dorsal lateral nucleus of the medial thalamus (DLM), comprise an analog to the human cortico-basal ganglia circuit. Intriguingly, lesions to these areas have massive, but contrary, effects on song learning in juvenile zebra finches. While lesions to Area X result in song elongation and high note-level variability in sequences, LMAN lesions cause an abrupt, premature crystallization of a highly repetitious single note song (Scharff and Nottebohm, 1991). Together, these results suggest an important contribution of non-auditory structures, specifically the basal ganglia, in serial order learning of natural communication sounds.

"WHAT THEN DETERMINES THE ORDER?"

Serial expertise serves an undoubtedly fundamental role in any system evaluating the order of temporal signals. Behaviorally, it seems that much is shared in the manner by which non-human animals and human adults learn and encode serially arranged stimuli (for review, see Terrace, 2011; Miller and Bee, 2012). Here, we have provided some accounts of how complex, temporally organized signals are encoded in the songbird telencephalon at the single neuron level, primarily focusing on the auditory pathways. Special attention has been paid to the increasing timescale of order representations as one ascends from the primary to secondary auditory areas. In humans, serial order representation in language is likely to operate in a similarly hierarchical fashion to integrate sequential information across different timescales (i.e., phoneme to morpheme, morpheme to phrase). And, while fMRI research has begun to localize major regions of interest where serial order encoding might take place (Henson et al., 2000; Marshuetz et al., 2000), non-human animal work certainly has much to contribute regarding what might be occurring within these areas at a much finer temporal resolution. Thus, while much remains unknown

regarding how the aforementioned neural mechanisms in songbirds scale up in the human brain to help support the kinds of complex temporal processing characteristic of language, we believe a comparative approach to be a necessary step toward this end.

CATEGORIZATION

SPEECH CATEGORIZATION IN HUMANS

At this stage, we have considered a system that successfully segregates and orders objects from an auditory stream. Now, we focus on a second cognitive ability necessary for language: categorization. When listening to someone speak, humans must link acoustic information to linguistic representations. However, this process is no simple feat. Consider the challenge of mapping a component of the speech signal to the most elementary linguistic unit: the phoneme. First, phonemes are high-dimensional acoustic objects, where features such as voice onset time, formant frequency, fundamental frequency and others all contribute to their descriptions. There is no single parameter dictating the boundary between all phoneme categories. Second, the acoustics of these phonemes are not static when repeated. Thus, a considerable challenge for the auditory system is correctly mapping phonetic utterances to representations in the face of considerable heterogeneity in their spectral and temporal characteristics across renditions and speakers.

The notion of categorical perception in speech was classically demonstrated by the work of Liberman and colleagues (reviewed in Liberman et al., 1967). Here, individuals were presented with a variety of speech-like sounds incrementally changed along a phonetically informative acoustic feature (for these stimuli, the slope of the rising transition between the first and second formant) and asked to denote if the sound they heard was *a/b/*, */d/*, or */g/*. Though the stimulus incrementally changed along a continuum, subjects' report of the phoneme heard did not follow a similar trajectory. Instead, there were pronounced boundaries between phonemes. Additionally, discrimination performance was significantly improved for stimulus pairs that were between phonemes opposed to those that were within a single phoneme class (Liberman et al., 1967).

Establishing categorical phoneme boundaries is experience-dependent with sounds placed into language-specific functional groups during early development (Holt and Lotto, 2008). In the most canonical demonstration of learned categories in speech perception, Miyawaki et al. (1975) tested the discrimination abilities of Japanese and American adults with */r/* and */l/* phonemes. Using a similar procedure to that of Liberman et al. (1967), American subjects showed a strong performance of between-phoneme discrimination and poor within-phoneme discrimination. Japanese subjects, on the other hand, maintained chance discrimination levels across all stimulus pairs. Importantly, the difference in discrimination performance between American and Japanese adults in this study maps onto the use of these phonemes in their native languages, thereby supporting the idea of learned phoneme categories.

Developmental studies support these results by showing a perceptual reorganization for phoneme contrasts during an infant's first year of life. For instance, 6- to 8-month-old English infants discriminate Hindi syllables not found in their parents' native language, but can no longer do so at 11–13 months. Those Hindi

syllables that are shared with English, however, remained highly discriminable across both age groups (Werker and Lalonde, 1988). But what is the mechanism underlying this perceptual learning of categorical boundaries? One leading hypothesis suggests that infants rapidly assess distributional patterns of auditory stimuli to determine categories. To test this idea, Maye et al. (2002) exposed 6- to 8-month-old infants to speech sounds varied along a phonetic continuum. For half of these subjects, the frequency distribution of sounds along this continuum was unimodal with a peak occurring in the center of the continuum. For the other half of subjects, the distribution was bimodal with the peaks occurring at opposing ends of the continuum. After just 2-min of exposure, the authors employed a looking time procedure to assess the infants' discrimination of the stimuli at the endpoints of the continuum. While infants from the bimodal exposure successfully discriminated sounds during this test phase, subjects from exposed to phonemes from a unimodal distribution failed to do so. This demonstrates that young infants spontaneously utilize distributional information from auditory stimuli to determine categorical boundaries.

CATEGORIZATION IN ANIMALS

The perceptual tools for phoneme categorization were argued to represent unique aspects of human language, unavailable to non-human animals (Liberman et al., 1967). Kuhl and Miller (1975) demonstrated the falsity of this claim by training chinchillas with an avoidance conditioning procedure, using */d/* and */t/* consonant-vowel syllables (from many human speakers) as stimuli. Chinchillas succeeded on this task and were able to generalize knowledge of */d/* and */t/* to utterances from totally novel speakers as well as to instances with novel vowel pairings (e.g., */da/* and */ta/*). The acoustic feature most indicative of the difference between this phoneme pair is the voice onset time (VOT). Thus, a VOT of 0 ms is readily perceived as */d/* in humans and an 80-ms VOT as */t/*. Chinchillas receiving feedback on these two stimuli demonstrate nearly identical psychophysical functions as humans when tested on VOTs between 0 and 80 ms. This study debunked the popular notion that humans possessed species-specific mechanisms necessary to perceive phonemes. Later work in chinchillas and monkeys further demonstrated learning of perceptual categories along single stimulus dimensions, such as VOT and formant spacing.

Evidence of human phoneme categorization in birds was demonstrated by Kluender et al. (1987), who showed that Japanese Quail can categorize three-phoneme syllables (consonant–vowel–consonant) according to the phonetic categories of the initial voiced stop consonants */d/*, */b/* and */g/*. Not only did this work expand the range of species showing categorical learning, but it demonstrated an arguably more advanced form of categorization, wherein the categorical discrimination must be performed along multiple stimulus dimensions (that is, not exclusively VOT or formant spacing). Categorization across variant phonemes was extended in a recent study (Ohms et al., 2010) that presented sounds in spoken words to zebra finches. The finches learned to discriminate minimal pairs (i.e., varying by a single phonological item, such as vowel; *wit* versus *wet*) and maintained performance across novel speakers and genders. Consistent with prior studies in humans, the results suggest that the finches used information about formant patterns.

In addition to songbirds ability to discriminate and categorize human vocalizations, they show expertise at categorizing their own vocalizations as well. Swamp sparrows exhibit categorical perception of notes of swamp sparrow song based on note duration (Nelson and Marler, 1989). This categorical boundary varies among different populations of swamp sparrows, reflecting an experience-dependent dialect.

NEUROBIOLOGY OF AUDITORY CATEGORIZATION IN SONGBIRDS

Not only do songbirds show categorical learning both human and songbird vocalizations, but work over the past decade has begun to elucidate correlated changes in neural coding throughout the auditory forebrain. Though categorical representations of sensory input have been studied extensively through human imaging and rodent and primate electrophysiology, our discussion here will focus on the evidence for categorical processing in avian neural systems, referencing mammalian work where appropriate.

Within the sensorimotor nucleus HVC of swamp sparrows, X-projecting neurons (HVCx) show categorical selectivity for note duration (Prather et al., 2009). These responses match the experience-dependent perceptual boundary such that cells show stronger responses for notes of within-category durations. The interneurons that provide input to HVCx neurons do not, however, show a categorical response boundary. Though the network connectivity and computational mechanisms giving rise to this categorical neuronal response is not clear, the implication that this temporally sensitive computation is taking place within HVC offers an opportunity to explore the microcircuit mechanism underlying categorical perception.

In addition to establishing perceptual categories along single dimensions, language also requires categorization of complex objects where the categorization does not necessarily reflect physical relationships between objects. Songbirds must make similar categorical discriminations in order to identify other individuals. Interestingly, HVC is also implicated in this type of non-vocal perceptual discrimination and may have a specific role in associating classes of stimuli to appropriate behavioral responses (Gentner et al., 2000).

How are learned categories represented by individual neurons and larger populations of neuronal activity? Where along the hierarchy of auditory processing afferent to HVC do these computations take place? After training starlings to discriminate sets of songs via operant conditioning, the responses of cells in CLM and CMM reflect these categories. In CMM, differences between learned categories can be observed in the average responses across cells to different learned categories, based on the reward associated with items of each class (Gentner and Margoliash, 2003). However, when both classes of stimuli are associated with reward, there is an increase in the firing rate of CMM neurons to both classes. Nonetheless, though mean firing rates increase in both nuclei, the variance of responses changes such that information about learned categories is encoded in the firing rates of individual cells in both CMM and CLM though there is more information about learned categories in CMM (Jeanne et al., 2011). The increased category information encoding for CMM neurons is accomplished through increasing the variance in differences between the firing rates to the different categories. These results bear some resemblance to

work implicating primate secondary auditory cortex in categorical representations (Tsunada et al., 2011). It will be important to develop careful behavioral paradigms to further elucidate the extent to which these response differences reflect learning about differential reward contingencies, associated behavioral responses, or stimulus class independent of associations. Such behavioral paradigms will enable to parsing apart the extent to which learning effects in CLM and CMM are driven by familiarization or behavioral association and explore how this information flows between these critical regions. It is not clear how these higher auditory areas interact to establish categories, but a picture is beginning to emerge whereby category representations are shaped by both single neuron and population processes interacting across a number of higher auditory regions. Further, chronic recording techniques will offer us the opportunity to explore how category representation in single neurons and populations are established during the learning of stimulus categories and the extent to which secondary auditory representations contribute to the animal's perception of stimuli.

RELATIONAL ABSTRACTION

RELATIONAL ABSTRACTION IN HUMANS

The ability to apprehend and generalize relationships between perceptual events is a fundamental component of human cognition and a crucial capacity for language comprehension and production. As such, we will not spend time discussing human capacity for relational abstraction here and will instead focus on non-human animal performance with a special focus on avian cognition (for review of human capacities for relational abstraction, see Hauser et al., 2002).

RELATIONAL ABSTRACTION IN ANIMALS

Non-human animals, too, are able to solve discriminations on the basis of relational information although there may be relevant limits to these abilities in non-humans. Traditionally, researchers have studied the abilities of non-human animals to learn about relationships between perceptual events using "match-to-sample" (MTS) tasks in which the animal is presented with a sample stimulus and then a target stimulus that either matches or does not match the sample. The animal is trained through successive reinforcement to give one response when the sample and target match and another when they do not. The interesting test then comes when the animal is presented with novel matching and non-matching stimuli. If the animal has learned to respond based on the relationship between the sample and target, then changing their physical properties should not effect responding. Many species of animals show exactly this kind of generalization to novel stimuli, including chimpanzees (Oden et al., 1988), monkeys (Mishkin et al., 1962), dolphins (Herman and Gordon, 1974), sea lions (Kastak and Schusterman, 1994), corvids (Wilson et al., 1985), and pigeons (Wasserman et al., 1995).

More stringent tests of relational abstraction, referred to as relational match-to-sample (RMTS) tasks, have also been devised in which animals are required to make judgments about relations between relations (Premack, 1983). In the original versions of these task, chimpanzees were presented with a pair of identical objects, "A A," or non-matching objects, "A B," as a sample stimulus, and

then had to choose among a second set of target objects that again either matched, “C C,” or not “C D.” If the sample pair matched, the chimps’ task was to choose the matching target pair, etc. This task is quite difficult for chimps to learn (Premack, 1983; see also, Thompson and Oden, 2000) and young children have difficulty with these kinds of tasks as well (House et al., 1974). Nonetheless, it is now clear that many species can learn to solve RMTS task, including chimpanzees (Premack, 1983; Thompson et al., 1997) and other apes (Vonk, 2003), parrots (Pepperberg, 1987), dolphins (Herman et al., 1993; Mercado et al., 2000), baboons (Bovet and Vauclair, 2001; Fagot et al., 2001), and pigeons (Blaisdell and Cook, 2005; Katz and Wright, 2006).

Exactly how animals learn to solve MTS and RMTS task is the subject of considerable debate and many researchers have noted that both tasks may reduce to classifications along perceptual dimensions of the stimuli (see Wasserman et al., 2004 for a partial review). Thus, while both MTS and RMTS performance require the abstraction of one or more “rules,” neither task may require the abstraction of a stimulus-independent, generalizable, and explicit concept of “sameness” (see Penn et al., 2008, for excellent discussion of these ideas). Even if relational abstraction is limited in animals in ways that it is not in humans, however, it nonetheless remains true that animals are able to learn novel and sophisticated, rule-governed behaviors that generalize beyond explicit perceptual cues but which are tied to specific dimensions of stimulus control (Penn et al., 2008). Appreciating the kind of rule-based, but stimulus-controlled, generalization behavior highlighted by the MTS and RMTS literature is helpful in understanding other kinds of pattern recognition in animals, particularly AGL.

Identifying patterns of events essential for adaptive behaviors like communication is a challenge shared across many taxa. In humans, recognizing patterned relationships between sound elements in acoustic streams plays an important role in many aspects of language acquisition. Beyond statistical expertise (as described above), humans learn the underlying rules to which patterned sequences ascribe. Indeed, infants at 7-months of age generalize phoneme-sequence patterns to novel strings of phonemes, mechanisms arguably necessary, though insufficient, for attaining knowledge of language (Marcus et al., 1999, 2007). Despite the importance of these capacities in human cognition, less is known of their function in non-human animals. Thus, our understanding of the biological substrate of temporal pattern and rule learning remains impoverished.

The capacity of non-human animals to learn temporal patterns has been demonstrated in several species, including songbirds (Gentner et al., 2006; Abe and Watanabe, 2011), chimpanzees (Endress et al., 2010), rhesus macaques (Hauser and Glynn, 2009), cotton-top tamarins (Versace et al., 2008; Endress et al., 2009a), rats (Murphy et al., 2004), pigeons (Herbranson and Shimp, 2003, 2008), and human infants (Reber, 1969; Gomez and Gerken, 1999; Marcus et al., 1999, 2007). Although several of these studies examined (and showed strong evidence for) generalization to novel patterns following the learned rules, few have systematically explored whether animals are capable of generalizing the acquired patterns to sequences built from entirely novel elements. For example, Starlings can learn to classify sequences of acoustically complex natural vocal (song) units, called “motifs,” whose

patterning is defined by at least two different grammatical forms: A^2B^2 and $(AB)^2$ (i.e., AABB and ABAB; Gentner et al., 2006). Here, “A” and “B” denote sets of acoustically distinct “rattle” and “warble” motifs, respectively. Having learned these patterns, they can generalize the acquired knowledge to novel sequences drawn from the same language, i.e. those constructed using the same A^2B^2 and $(AB)^2$ patterning rules using the same set of A and B motifs heard in training (Gentner et al., 2006). It is unclear, however, the extent to which knowledge of the acquired patterns might exist independent of the constituent stimuli, and thus whether it would generalize to novel warbles and rattles or other sequences of defined acoustic perceptual categories.

A recent investigation by Murphy et al. (2008) addresses some of these questions, exploring the rule learning abilities of rats. The experimenters asked whether rats could learn a pattern of events predicting the forthcoming availability of food in a Pavlovian conditioning task. In the first of two experiments, food was administered following one of three rule-governed sequences of bright and dim light presentations, such as the form XYX . In this example, food was presented following either bright–dim–bright or dim–bright–dim light cycles, whereas light changes of the form XXY or YXX were not associated with food. Rats exhibited greater anticipatory behavior in the wake of XYX patterns, suggesting they learned which pattern preceded food delivery (though see Toro and Trobalón, 2005, for failure of rats to learn these patterns built from human phonemes in a lever-pressing task). In a second experiment, researchers asked whether this pattern knowledge could be transferred to novel events following the same patterning rule. Here, subjects were presented with tone sequences governed by the same temporal structure as those used in experiment one. For example, a subject might be trained that the tone sequences 3.2–9–3.2 or 9–3.2–9 kHz (again, an XYX pattern) indicate the imminent availability of food. After once more learning which was the predictive pattern, a non-overlapping distribution of tones arranged in the various patterns were presented (e.g., 12.5–17.5–12.5 or 17.5–12.5–17.5 kHz for XYX , versus XXY or YXX). Under these conditions, rats were able to transfer rule knowledge from one sequence of tones to a novel sequence sharing the same abstracted relationship.

The results of AGL tasks extend the evidence for rule learning and generalization into the temporal domain. One advantage of these kinds of task is that they can easily be adapted to incorporate increasingly complex stimulus constructions that capture theoretically more complex rules. For example, Gentner et al. (2006) demonstrated that both a finite state and a context-free patterning rule could be learned by starlings. More recently, studies extended this work to show that Bengalese finches, another songbird with syntactically variable songs, could recognize violations to artificial grammar containing center-embedded structures (Abe and Watanabe, 2011). To create this grammar, the researchers defined three classes A, C, and F, consisting of four Bengalese finch song syllables each. The four A syllables were each matched with a particular F syllable. Interposed between them was a “C phrase” which was either any C syllable, or another matched AF pair. During exposure, the Bengalese finches heard every possible grammatical string consisting of ACF, and about half of the possible $AA'CF'F$ stimuli. During testing, the finches heard novel grammatical $AA'CF'F$

strings, as well as sequences that were ungrammatical. Shifts in their call rates to the test strings were then used as evidence that the finches detected a difference between the test stimuli and their habituated grammatical stimuli. These results revealed a striking sensitivity to the recursive structure of the grammatical strings the finches were exposed to – an important result.

There remain, however, many significant questions about kinds of rules that animals acquire in each of these different AGL training conditions (see van Heijningen et al., 2009; Beckers et al., 2012), and of how such rules relate to underlying perceptual dimensions of the stimuli. In many cases, the encoding strategies employed to generalize the rule have not been fully explored. Ascertaining such strategies is crucial to understanding of the types of temporal information amenable to generalization and to future understanding of the underlying neurobiology. While many non-human animals display abilities to learn and generalize temporal pattern rules, it may be that their abilities to abstract such rules beyond the perceptual dimensions of the constituent stimuli are quite limited. Thus the uniqueness of human syntax may lie not in its computational sophistication, but rather in its independent representation and use these patterning rules at levels of abstraction far removed from the specific speech (or manual gesture) signals. If true, then understanding the neurobiology of rule abstraction, in any context, will be crucial to understanding the neurobiology of language.

NEUROBIOLOGY OF RELATIONAL ABSTRACTION

Initial investigations of the neural substrates for MTS behaviors were concerned with memory processes rather than relational abstraction. Such studies led, in any case, to the remarkable finding of individual neurons in the prefrontal cortex (PFC) that show sustained activity during the interval between the presentation of the initial sample and later target stimulus (Fuster and Alexander, 1971). These responses have been broadly interpreted as memory traces for the physical attributes of the sample (or prospective coding of the target) stimulus. Indeed responses in many of these neurons are preferential for specific sensory domains and track physical dimensions of the stimuli within them (see Fuster, 2009 for review). Regions of the PFC have long been associated, based on lesion evidence in humans (Milner, 1982), with impairments in the ability to flexibly change rules for classifying simple visual stimuli – typically cards that can be sorted differentially according to several different features. More recent work indicates that neural correlates to “simple” rules like shape are strong in the PFC of both monkeys and rats (Hoshi et al., 1998; Asaad et al., 2000; Schoenbaum and Setlow, 2001; Wallis et al., 2001). All these data are consistent with the idea the PFC is involved in abstracting sensory information across many domains reflecting either working memory or more explicit representations of rules.

One insight into this function comes from PFC neurons recorded while monkeys alternated between “same” and “different” responses on a MTS task. When responding correctly to novel stimuli (pictures) roughly 40% of the neurons in PFC show firing rate changes that reflected the rule the monkey was currently using. Moreover, like the rule itself these neuronal responses generalized across different cues used to signal the rule, and were not linked to the behavioral response (Wallis and Miller, 2003). Similar pattern

of PFC activity are observed in imaging data from humans when they are retrieving or maintaining abstract rules (Bunge et al., 2003). Mapping regions in the monkey onto the human brain is no simple matter, but at least a subset of the regions identified through these and many other similar single neuron studies (see Miller, 2007 for recent review) may directly correspond to human frontal regions thought, from a large corpus of imaging work, to underlie more direct language behavior (reviewed in Friederici, 2011).

One limitation of the current work in both monkeys and human is that it remains very correlative and focused on localization rather than the underlying neural mechanisms and computations that might support relational abstraction and rule encoding. More mechanistic studies will require training non-human primates on increasingly complicated rule abstraction tasks that have direct ties to language processing, but this has been difficult (Miller, 2007). Alternatively, recording in PFC-like regions in songbirds (Güntürkün, 2011) trained on complex syntactic processing tasks will be equally useful.

CONCLUSION: TOWARD A COMPARATIVE NEUROBIOLOGY OF LANGUAGE

The large suite of behaviors encapsulated by language constitutes (arguably) the most complex set of cognitive capacities that neurobiology can attempt to explain. This endeavor presents substantial challenges. Neurobiology has been very good at characterizing the role of neurons and populations of neurons in visual perception – the dominant field for sensory driven cognition. Yet, language is at its core a temporally dynamic process, emerging over the timespan of syllables, words, and sentences. As soon as new information is acquired, other parts of the signal are gone. The continuous stream of auditory information must be segmented, individual elements categorized. New information must be processed and compared with recent words or words long past. Different levels of cognitive processes must interact as new signals force the reevaluation of earlier computations – one word might alter the entire grammatical structure and the meaning of a sentence. Language perception (and production), therefore, necessitate novel fundamental neurobiological mechanisms that can accommodate these rapid temporal dynamics that vision neurobiology has simply not delivered.

To an even greater extent, neurobiology presents a difficult problem for linguistics. Biology imposes a number of constraints on our understanding of cognitive processes and limits the plausibility of cognitive models that remain agnostic to biological instantiations. The inconsistencies between theoretical and behavioral linguistics are longstanding, but rather than conceptualizing biological processes such as working memory as “constraints” on an otherwise perfect computing system without resource limitations, we should recognize that the neurobiology is precisely what enables these computations in the first place. The challenge for linguistics is determining if and how theoretical linguistic work is instantiated in biology, not the other way around. Attempting to pursue a research agenda to understand how language evolved while ignoring biology is a fool’s errand. To fully understand the evolution of language, we need a research program firmly rooted in understanding the underlying neurobiology.

One area where significant effort is urgently required is the development of biologically plausible, neuron level, computational models for cognitive components of language such as grammatical processing. For those computational models that show moderate success at replicating human grammatical processing (such as Simple Recurrent Networks), it is unknown whether the biological network architectures they require actually exist. Do such architectures exist in nature and can they account for natural processing of grammars? If not, what are the architectures that support these computations? What sub-populations of neurons are involved? How are they wired together locally and between systems? How do the time-varying dynamics of excitatory and inhibitory neural processes contribute to these computations? These and similar questions cannot be addressed with current methods in human neuroscience. Though natural lesion, functional imaging, and electrophysiology studies in humans offer insights into the functional anatomy and large-scale dynamics of language, understanding of neuronal-level processing requires more direct measurement and manipulation of the neurons that make up the brain.

We have proposed that songbirds offer an opportunity to explore how individual and populations of neurons contribute to at least some of the cognitive processes that are requisite to language. Songbirds are one of the few classes of animals that exhibit vocal learning, and are already an established modern system for studying these complex communication behaviors. Exploiting the neurobiology of song perception in these animals holds the promise of a nearly complete animal model for learned vocal communication. In the present review we have attempted to highlight several areas that we think are both tractable for neurobiological study in the immediate future and directly relevant to language in humans. We think that this effort will inform a number of specific debates in human language perception. For example, what is the role of the basal ganglia and motor systems in speech and

language perception? The implication in grammatical processing of a basal ganglia nucleus known to be involved in vocal production (Abe and Watanabe, 2011) is an interesting result and further research could shed some light on how motor systems support serial expertise and relational abstraction. Likewise, how valid are “dual stream” models of language processing that can either blur or dissociate auditory comprehension from auditory–motor interaction (Hickok and Poeppel, 2007)? The dual stream processing in vision has been strongly supported by primate neurobiology (Goodale and Milner, 1992; Milner and Goodale, 2008), but the auditory domain would benefit from a comprehensive model for perception and production.

Eventually, we hope that continued improvements in non-invasive neural recording methods will instantiate a complete neurobiology of language in humans, and we are open to the possibility that there may well be unique computational or physiological features of language perception for which no appropriate animal model exists. But to the best of our present knowledge, vertebrate nervous systems, and the kinds of computations, network circuitry, and dynamics they employ are remarkably similar. If we 1 day get to the point where we understand enough about the neural mechanisms of these requisite processes of language to know why such animal models are no longer useful, we will have achieved quite a lot. Until then, there is no doubt that comparative approaches still have much to contribute to our understanding of language.

ACKNOWLEDGMENTS

This material is based upon work supported by a National Science Foundation Graduate Research Fellowship under Grant No. 2009060397 to Justin T. Kiggins, a National Science Foundation Graduate Research Fellowship under Grant No. 2011122846 to Jordan A. Comins, and NIH DC008358 to Timothy Q. Gentner. Two anonymous reviewers provided valuable comments on an earlier draft of the manuscript.

REFERENCES

- Abe, K., and Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* 14, 1067–1074.
- Abla, D., Katahira, K., and Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *J. Cogn. Neurosci.* 20, 952–964.
- Asaad, W. F., Rainer, G., and Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *J. Neurophysiol.* 84, 451–459.
- Balleine, B. W., Garner, C., Gonzalez, E., and Dickinson, A. (1995). Motivational control of heterogeneous instrumental chains. *J. Exp. Psychol. Anim. Behav. Process.* 21, 203–217.
- Beckers, G. J. L., Bolhuis, J. J., Okanoya, K., and Berwick, R. C. (2012). Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* 23, 139–145.
- Berwick, R., Okanoya, K., Beckers, G., and Bolhuis, J. (2011). Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci. (Regul. Ed.)*
- Blaisdell, A. P., and Cook, R. G. (2005). Two-item same-different concept learning in pigeons. *Anim. Learn. Behav.* 33, 67–77.
- Bovet, D., and Vauclair, J. (2001). Judgment of conceptual identity in monkeys. *Psychon. Bull. Rev.* 8, 470–475.
- Brainard, M. S., and Doupe, A. J. (2002). What songbirds teach us about learning. *Nature* 417, 351–358.
- Bunge, S. A., Kahn, I., Wallis, J. D., Miller, E. K., and Wagner, A. D. (2003). Neural circuits subserving the retrieval and maintenance of abstract rules. *J. Neurophysiol.* 90, 3419–3428.
- Chen, S., Swartz, K. B., and Terrace, H. S. (1997). Knowledge of the original position of list items in rhesus monkeys. *Psychol. Sci.* 8, 80–86.
- Comins, J. A., and Gentner, T. Q. (2010). Working memory for patterned sequences of auditory objects in a songbird. *Cognition* 117, 38–53.
- D’Amato, M. R., and Colombo, M. (1988). Representation of serial order in monkeys (*Cebus apella*). *J. Exp. Psychol. Anim. Behav. Process.* 14, 131–139.
- Drew, P. J., and Abbott, L. F. (2003). Model of song selectivity and sequence generation in area HVC of the songbird. *J. Neurophysiol.* 89, 2697–2706.
- Eens, M., Pinxten, R., and Verheyen, R. F. (1988). Temporal and sequential organisation of song bouts in the starling. *Ardea* 77, 75–86.
- Endress, A. D., Cahill, D., Block, S., Watumull, J., and Hauser, M. D. (2009a). Evidence of an evolutionary precursor to human language affixation in a non-human primate. *Biol. Lett.* 5, 749–751.
- Endress, A. D., Nespor, M., and Mehler, J. (2009b). Perceptual and memory constraints on language acquisition. *Trends Cogn. Sci. (Regul. Ed.)* 13, 348–353.
- Endress, A. D., Carden, S., Versace, E., and Hauser, M. D. (2010). The apes’ edge: positional learning in chimpanzees and humans. *Anim. Cogn.* 13, 483–495.
- Endress, A. D., and Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cogn. Psychol.* 61, 177–199.
- Endress, A. D., Scholl, B. J., and Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *J. Exp. Psychol. Gen.* 134, 406.
- Fagot, J., Wasserman, E. A., and Young, M. E. (2001). Discriminating the relation between relations: the role of entropy in abstract conceptualization by baboons (*Papio papio*) and humans (*Homo sapiens*). *J. Exp. Psychol. Anim. Behav. Process.* 27, 316–328.

- Fitch, W. T., and Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.
- Fitch, W. T., Hauser, M. D., and Chomsky, N. (2005). The evolution of the language faculty: clarifications and implications. *Cognition* 97, 179–210; discussion 211–225.
- Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392.
- Friederici, A. D., Friedrich, M., and Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Curr. Biol.* 17, 1208–1211.
- Fuster, J. M. (2009). Cortex and memory: emergence of a new paradigm. *J. Cogn. Neurosci.* 21, 2047–2072.
- Fuster, J. M., and Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science* 173, 652–654.
- Gentner, T. Q. (2008). Temporal scales of auditory objects underlying bird-song vocal recognition. *J. Acoust. Soc. Am.* 124, 1350–1359.
- Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.
- Gentner, T. Q., and Hulse, S. H. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Anim. Behav.* 56, 579–594.
- Gentner, T. Q., Hulse, S. H., Bentley, G. E., and Ball, G. F. (2000). Individual vocal recognition and the effect of partial lesions to HVC on discrimination, learning, and categorization of conspecific song in adult songbirds. *J. Neurobiol.* 42, 117–133.
- Gentner, T. Q., Hulse, S. H., Duffy, D., and Ball, G. F. (2001). Response biases in auditory forebrain regions of female songbirds following exposure to sexually relevant variation in male song. *J. Neurobiol.* 46, 48–58.
- Gentner, T. Q., and Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature* 424, 669–674.
- Gill, P., Woolley, S. M. N., Fremouw, T., and Theunissen, F. E. (2008). What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. *J. Neurophysiol.* 99, 2809–2820.
- Gomez, R. L., and Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70, 109–135.
- Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25.
- Güntürkün, O. (2011). The convergent evolution of neural substrates for cognition. *Psychol. Res.* 76, 212–219.
- Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Hauser, M. D., and Glynn, D. (2009). Can free-ranging rhesus monkeys (*Macaca mulatta*) extract artificially created rules comprised of natural vocalizations? *J. Comp. Psychol.* 123, 161–167.
- Henson, R., Burgess, N., and Frith, C. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia* 38, 426–440.
- Henson, R. N. A., Norris, D. G., Page, M. P. A., and Baddeley, A. D. (1996). Unchained memory: error patterns rule out chaining models of immediate serial recall. *Q. J. Exp. Psychol. A* 49, 80–115.
- Herbranson, W. T., and Shimp, C. P. (2003). “Artificial grammar learning” in pigeons: a preliminary analysis. *Anim. Learn. Behav.* 31, 98–106.
- Herbranson, W. T., and Shimp, C. P. (2008). Artificial grammar learning in pigeons. *Learn. Behav.* 36, 116–137.
- Herman, L. M., and Gordon, J. A. (1974). Auditory delayed matching in the bottlenose dolphin. *J. Exp. Anal. Behav.* 21, 19–26.
- Herman, L. M., Kuczaj, S. A., and Holder, M. D. (1993). Responses to anomalous gestural sequences by a language-trained dolphin: evidence for processing of semantic relations and syntactic information. *J. Exp. Psychol. Gen.* 122, 184–194.
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Holt, L. L., and Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Curr. Dir. Psychol. Sci.* 17, 42–46.
- Hoshi, E., Shima, K., and Tanji, J. (1998). Task-dependent selectivity of movement-related neuronal activity in the primate prefrontal cortex. *J. Neurophysiol.* 80, 3392–3397.
- House, B. J., Brown, A. L., and Scott, M. S. (1974). Children's discrimination learning based on identity or difference. *Adv. Child. Dev. Behav.* 9, 1–45.
- Iverson, J. R., Patel, A. D., and Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *J. Acoust. Soc. Am.* 124, 2263–2271.
- Jackendoff, R., and Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition* 97, 211–225.
- Jeanne, J. M., Thompson, J. V., Sharpee, T. O., and Gentner, T. Q. (2011). Emergence of learned categorical representations within an auditory forebrain circuit. *J. Neurosci.* 31, 2595–2606.
- Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends Cogn. Sci. (Regul. Ed.)* 3, 323–328.
- Kastak, D., and Schusterman, R. J. (1994). Transfer of visual identity matching-to-sample in two California sea lions (*Zalophus californianus*). *Anim. Learn. Behav.* 22, 427–435.
- Katz, J. S., and Wright, A. A. (2006). Same/different abstract-concept learning by pigeons. *J. Exp. Psychol. Anim. Behav. Process.* 32, 80–86.
- Kelley, D. B., and Nottebohm, F. (1979). Projections of a telencephalic auditory nucleus-field L in the canary. *J. Comp. Neurol.* 183, 455–469.
- Kluender, K. R., Diehl, R. L., and Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science* 237, 1195–1197.
- Kuhl, P., and Miller, J. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* 190, 69–72.
- Lashley, K. (1951). “The problem of serial order in behavior,” in *Cerebral Mechanisms in Behavior*, Vol. 26, ed. L. A. Jeffress (New York: Wiley), 112–137.
- Lewicki, M. S., and Arthur, B. J. (1996). Hierarchical organization of auditory temporal context sensitivity. *J. Neurosci.* 16, 6987–6998.
- Lewicki, M. S., and Konishi, M. (1995). Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc. Natl. Acad. Sci. U.S.A.* 92, 5582–5586.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431.
- Lipkind, D., and Tchernichovski, O. (2011). Quantification of developmental birdsong learning from the subsyllabic scale to cultural evolution. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 3), 15572–15579.
- Marcus, G., Vijayan, S., and Rao, S. B. (1999). Rule learning by seven-month-old infants. *Science* 283, 77–80.
- Marcus, G. F., Fernandes, K. J., and Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychol. Sci.* 18, 387–391.
- Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.* 3, 1039–1057.
- Margoliash, D. (1986). Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J. Neurosci.* 6, 1643–1661.
- Margoliash, D., and Fortune, S. (1992). Temporal and harmonic combination-sensitive finch's HVC neurons in the zebra. *J. Neurosci.* 12, 4309–4326.
- Margoliash, D., and Nusbaum, H. C. (2009). Language: the perspective from organismal biology. *Trends Cogn. Sci. (Regul. Ed.)* 13, 505–510.
- Marshuetz, C., Smith, E. E., Jonides, J., Degutis, J., and Chenevert, T. L. (2000). Order information in working memory: fMRI evidence for parietal and prefrontal mechanisms. *J. Cogn. Neurosci.* 12, 130–144.
- Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111.
- McNealy, K., Mazziotta, J. C., and Dapretto, M. (2010). The neural basis of speech parsing in children and adults. *Dev. Sci.* 13, 385–406.
- Meliza, C. D., Chi, Z., and Margoliash, D. (2010). Representations of conspecific song by starling secondary forebrain auditory neurons: toward a hierarchical framework. *J. Neurophysiol.* 103, 1195–1208.
- Mello, C. V., and Ribeiro, S. (1998). ZENK protein regulation by song in the brain of songbirds. *J. Comp. Neurol.* 393, 426–438.
- Mercado, E., Killebrew, D. A., Pack, A. A., Mácha, I. V., and Herman, L. M. (2000). Generalization of “same-different” classification abilities in bottlenosed dolphins. *Behav. Processes* 50, 79–94.
- Miller, C. T., and Bee, M. A. (2012). Receiver psychology turns 20: is it time for a broader approach? *Anim. Behav.* 83, 331–343.
- Miller, E. (2007). “The prefrontal cortex: categories, concepts, and cognitive control,” in *Memories: Molecules and Circuits*, eds B. Bontempi, A. J. Silva, and Y. Christen (Heidelberg: Springer), 137–154.

- Milner, A. D., and Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia* 46, 774–85.
- Milner, B. (1982). Some cognitive effects of frontal-lobe lesions in man. *Philos. Trans. R. Soc. B Biol. Sci.* 298, 211–226.
- Mishkin, M., Prockop, E. S., and Rosvold, H. E. (1962). One-trial object-discrimination learning in monkeys with frontal lesions. *J. Comp. Physiol. Psychol.* 55, 178–181.
- Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., and Fujimura, O. (1975). An effect of linguistic experience—the discrimination of [r] and [l] by native speakers of Japanese and English. *Atten. Percept. Psychophys.* 18, 331–340.
- Murphy, R. A., Mondragón, E., and Murphy, V. A. (2008). Rule learning by rats. *Science* 319, 1849–1851.
- Murphy, R. A., Mondragón, E., Murphy, V. A., and Fouquet, N. (2004). Serial order of conditional stimuli as a discriminative cue for Pavlovian conditioning. *Behav. Processes* 67, 303–311.
- Nelson, D. A., and Marler, P. (1989). Categorical perception of a natural stimulus continuum: birdsong. *Science* 244, 976–978.
- Nishikawa, J., Okada, M., and Okanoya, K. (2008). Population coding of song element sequence in the Bengalese finch HVC. *Eur. J. Neurosci.* 27, 3273–3283.
- Oden, D. L., Thompson, R. K., and Premack, D. (1988). Spontaneous transfer of matching by infant chimpanzees (*Pan troglodytes*). *J. Exp. Psychol. Anim. Behav. Process.* 14, 140–145.
- Ohms, V. R., Gill, A., Van Heijningen, C. A. A., Beckers, G. J. L., and ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proc. Biol. Sci.* 277, 1003–1009.
- Penn, D. C., Holyoak, K. J., and Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31, 109–30; discussion 130–178.
- Pepperberg, I. M. (1987). Acquisition of the same/different concept by an African Grey parrot (*Psittacus erithacus*): learning with respect to categories of color, shape, and material. *Anim. Learn. Behav.* 15, 423–432.
- Pinker, S., and Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition* 95, 201–236.
- Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S., and Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nat. Neurosci.* 12, 221–228.
- Premack, D. (1983). The codes of man and beasts. *Behav. Brain Sci.* 6, 125–167.
- Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *J. Exp. Psychol.* 81, 115–119.
- Rose, G. J., Goller, F., Gritton, H. J., and Plamondon, S. L. (2004). Species-typical songs in white-crowned sparrows tutored with only phrase pairs. *Nature* 753–758.
- Saffran, J. J., Aslin, R. R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.
- Sanders, L. D., Newport, E. L., and Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nat. Neurosci.* 5, 700–703.
- Scharff, C., and 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. *J. Neurosci.* 11, 2896–2913.
- Schoenbaum, G., and Setlow, B. (2001). Integrating orbitofrontal cortex into prefrontal theory: common processing themes across species and subdivisions. *Learn. Mem.* 8, 134–147.
- Skinner, B. (1934). The extinction of chained reflexes. *Proc. Natl. Acad. Sci. U.S.A.* 20, 234–237.
- Sockman, K. W., Gentner, T. Q., and Ball, G. F. (2002). Recent experience modulates forebrain gene-expression in response to mate-choice cues in European starlings. *Proc. Biol. Sci.* 269, 2479–2485.
- Straub, R. O., and Terrace, H. S. (1982). Generalization of serial learning in the pigeon. *Anim. Learn. Behav.* 9, 454–468.
- Takahashi, M., Yamada, H., and Okanoya, K. (2010). Statistical and prosodic cues for song segmentation learning by bengalese finches (*Lonchura striata* var. domestica). *Ethology* 116, 481–489.
- Terrace, H. S. (2005). The simultaneous chain: a new approach to serial learning. *Trends Cogn. Sci. (Regul. Ed.)* 9, 202–210.
- Terrace, H. (2011). “Missing links in the evolution of language,” in *Characterizing Consciousness: From Cognition to the Clinic? Research and Perspectives in Neurosciences*, eds S. Dehaene and Y. Christen (Berlin: Springer-Verlag), 1–25.
- Theunissen, F. E., Sen, K., and Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.* 20, 2315–2331.
- Thiessen, E. D., and Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev. Psychol.* 39, 706–716.
- Thompson, J. V., and Gentner, T. Q. (2010). Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. *J. Neurophysiol.* 103, 1785–1797.
- Thompson, R. K., Oden, D. L., and Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *J. Exp. Psychol. Anim. Behav. Process.* 23, 31–43.
- Thompson, R. K. R., and Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: the paleological monkey and the analogical ape. *Cogn. Sci.* 24, 363–396.
- Toro, J. M., and Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Percept. Psychophys.* 67, 867–875.
- Tsunada, J., Lee, J. H., and Cohen, Y. E. (2011). Representation of speech categories in the primate auditory cortex. *J. Neurophysiol.* 105, 2634–2646.
- van Heijningen, C. A. A., de Visser, J., Zuidema, W., and ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20538–20543.
- Versace, E., Endress, A. D., and Hauser, M. D. (2008). Pattern recognition mediates flexible timing of vocalizations in nonhuman primates: experiments with cotton-top tamarins. *Anim. Behav.* 76, 1885–1892.
- Vonk, J. (2003). Gorilla (*Gorilla gorilla gorilla*) and orangutan (*Pongo abelii*) understanding of first- and second-order relations. *Anim. Cogn.* 6, 77–86.
- Wallis, J. D., Anderson, K. C., and Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature* 411, 953–956.
- Wallis, J. D., and Miller, E. K. (2003). From rule to response: neuronal processes in the premotor and prefrontal cortex. *J. Neurophysiol.* 90, 1790–1806.
- Washburn, M. F. (1916). *Movement and Mental Imagery: Outlines of a Motor Theory of the Complexer Mental Processes*. Boston: Houghton Mifflin Company.
- Wasserman, E. A., Hugart, J. A., and Kirkpatrick-Steger, K. (1995). Pigeons show same-different conceptualization after training with complex visual stimuli. *J. Exp. Psychol. Anim. Behav. Process.* 21, 248–252.
- Wasserman, E. A., Young, M. E., and Cook, R. G. (2004). Variability discrimination in humans and animals: implications for adaptive action. *Am. Psychol.* 59, 879–890.
- Watson, J. (1920). Is thinking merely the action of language mechanisms? *Br. J. Psychol.* 11, 87–104.
- Weisman, R. G., Wasserman, E. A., Dodd, P. W., and Larew, M. B. (1980). Representation and retention of two-event sequences in pigeons. *J. Exp. Psychol. Anim. Behav. Process.* 6, 312–325.
- Werker, J. F., and Lalonde, C. E. (1988). Cross-language speech perception: initial capabilities and developmental change. *Dev. Psychol.* 24, 672–683.
- Williams, H., and Staples, K. (1992). Syllable chunking in zebra finch (*Taeniopygia guttata*) song. *J. Comp. Psychol.* 106, 278–286.
- Wilson, B., Mackintosh, N. J., and Boakes, R. A. (1985). Transfer of relational rules in matching and oddity learning by pigeons and corvids. *Q. J. Exp. Psychol. B* 37, 313–332.
- Zokoll, M. A., Naue, N., Herrmann, C. S., and Langemann, U. (2008). Auditory memory: a comparison between humans and starlings. *Brain Res.* 1220, 33–46.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 December 2011; paper pending published: 11 January 2012; accepted: 21 March 2012; published online: 09 April 2012.

Citation: Kiggins JT, Comins JA and Gentner TQ (2012) Targets for a comparative neurobiology of language. *Front. Evol. Neurosci.* 4:6. doi: 10.3389/fnevo.2012.00006

Copyright © 2012 Kiggins, Comins and Gentner. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.