frontiers
RESEARCH TOPICS

NEUROBIOLOGY OF HUMAN LANGUAGE AND ITS EVOLUTION: PRIMATE AND NONPRIMATE PERSPECTIVES

Topic Editors
Constance Scharff, Angela D. Friederici
and Michael Petrides

frontiers in
EVOLUTIONARY NEUROSCIENCE

FRONTIERS COPYRIGHT STATEMENT

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-111-6
DOI 10.3389/978-2-88919-111-6

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.

All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

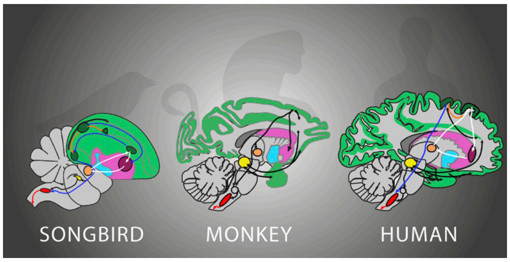
WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

NEUROBIOLOGY OF HUMAN LANGUAGE AND ITS EVOLUTION: PRIMATE AND NONPRIMATE PERSPECTIVES

Topic Editors:
Constance Scharff, Freie Universitaet Berlin, Germany
Angela D. Friederici, Max Planck Institute for Human Cognitive and Brain Sciences, Germany
Michael Petrides, Montreal Neurological Institute, McGill University, Canada



The evolution of human language has been discussed for centuries from different perspectives. Linguistic theory has proposed grammar as a core part of human language that has to be considered in this context. Recent advances in neurosciences have allowed us to take a new neurobiological look on the similarities and dissimilarities of cognitive capacities and their neural basis across both closely and distantly related species. A couple of decades ago the comparisons were mainly drawn between human and non-human primates, investigating the cytoarchitecture of particular brain areas and their structural connectivity. Moreover, comparative studies were conducted with respect to their ability to process grammars of different complexity. So far the available data suggest that non-human primates are able to learn simple probabilistic grammars, but not hierarchically structured complex grammars. The human brain, which easily learns both grammars, differs from the non-human brain (among others) in how two language-relevant brain regions (Broca's area and superior temporal cortex) are connected structurally. Whether the more dominant dorsal pathway in humans compared to non-human primates is causally related to this behavioral difference is an issue of current debate. Ontogenetic findings suggest at least a correlation between the maturation of the dorsal pathway and the behavior to process syntactically complex structures, although a causal prove is still not available. Thus the neural basis of complex grammar processing in humans remains to be defined.

More recently it has been reported that songbirds are also able to distinguish between sound sequences reflecting complex grammar. Interestingly, songbirds learn to sing by imitating adult song in a process not unlike language development in children. Moreover, the neural circuits supporting this behavior in songbirds bear anatomical and functional similarities to those in humans. In adult humans the fiber tract connecting the auditory cortex and motor cortex dorsally is known to be involved in the repetition of spoken language. This pathway is present already at birth and is taken to play a major role during language acquisition. In songbirds, detailed information exist concerning the interaction of auditory, motor and cortical-basal ganglia processing during song learning, and present a rich substrate for comparative studies.

The scope of the Research Topic is to bring together contributions of researchers from different fields, who investigate grammar processing in humans, non-human primates and songbirds with the aim to find answers to the question of what constitutes the neurobiological basis of grammar learning. Open questions are: Which brain networks are relevant for grammar learning? Is there more than one dorsal pathway (one from temporal cortex to motor cortex and one to Broca's area) and if so what are their functions? Has the ability to process sequences of a given hierarchical complexity evolved in different phylogenetic lines (birds, primates, other vocal production learners such as bats)? Is the presence of a sensory-to-motor circuit in humans a precondition for development of a dorsal pathway between the temporal cortex and Broca's area? What role do subcortical structures (Basal Ganglia) play in vocal and grammar learning?

Table of Contents

frontiers in
EVOLUTIONARY NEUROSCIENCE

EDITORIAL
published: 28 January 2013
doi: 10.3389/fnevo.2013.00001

Neurobiology of human language and its evolution: primate and non-primate perspectives

***Constance Scharff*[1], *Angela D. Friederici*[2]* and *Michael Petrides*[3]**

[1] *Departments of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Berlin, Germany*
[2] *Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[3] *Neurological Institute and Hospital, Montreal, QC, Canada*
**Correspondence: angelafr@cbs.mpg.de*

Edited by:
Ralph L. Holloway, Columbia University, USA

Reviewed by:
Ralph L. Holloway, Columbia University, USA

The evolution of human language has been discussed for centuries from different perspectives. Linguistic theory has proposed grammar as a core part of human language that has to be considered in this context. Recent advances in neurosciences have allowed us to take a new neurobiological look on the similarities and dissimilarities of cognitive capacities and their neural basis across both closely and distantly related species. A couple of decades ago, the comparisons were mainly drawn between human and non-human primates, investigating the cytoarchitecture of particular brain areas and their structural connectivity. Moreover, comparative studies were conducted with respect to their ability to process grammars of different complexity. So far the available data suggest that non-human primates are able to learn simple probabilistic grammars, but not hierarchically structured complex grammars. The human brain, which easily learns both grammars, differs from the non-human brain (among others) in how two language-relevant brain regions (Broca's area in the inferior frontal cortex and the superior temporal cortex) are connected structurally by fiber tracts which run dorsally and ventrally in the primate brain. Whether the more dominant dorsal pathway in humans compared to non-human primates is causally related to this behavioral difference is an issue of current debate. Ontogenetic findings suggest at least a correlation between the maturation of the dorsal pathway and the behavior to process syntactically complex structures, although the ultimate causal prove is still not available. Thus, the neural basis of complex grammar processing in humans remains to be defined.

More recently it has been reported that songbirds are also able to distinguish between sound sequences reflecting complex grammar. Interestingly, songbirds learn to sing by imitating adult song in a process not unlike language development in children. Moreover, the neural circuits supporting this behavior in songbirds bear anatomical and functional similarities to those in humans. In adult humans the fiber tract connecting the auditory cortex and motor cortex dorsally is known to be involved in the repetition of spoken language. This pathway is present already at birth and is taken to play a major role during language acquisition. In songbirds, detailed information exist concerning the interaction of auditory, motor, and cortical-basal ganglia processing during song learning, and present a rich substrate for comparative studies.

The scope of the Research Topic was to bring together contributions of researchers from different fields, who investigate grammar processing in humans, non-human primates, and songbirds with the aim to find answers to the question of what constitutes the neurobiological basis of language and language learning.

A number of contributions discuss the ventral and dorsal pathways in human and non-human primates considering their functional roles in speech and language. Some of these take an evolutionary perspective comparing non-human and human primates (Rauschecker, 2012; Rilling et al., 2012), whereas other takes an ontogenetic perspective (Friederici, 2012). The functional roles of the ventral and dorsal pathways in language and other modalities in particular action including articulatory and hand gestures are discussed in further articles (Fitch, 2011; Aboitiz, 2012; Rijntjes et al., 2012). Two articles consider the language system at the interface of two other human specific abilities, namely number processing (Heim et al., 2012) and reading (Lachmann et al., 2012). A couple of contributions take the evolutionary perspective even further by including song birds into their comparative approach (Berwick et al., 2012; Kiggins et al., 2012; Petkov and Jarvis, 2012).

The selection of the articles provides a picture of the current views on the evolutionary and neurobiological basis of the language and language learning.

REFERENCES

Aboitiz, F. (2012). Gestures, vocalizations, and memory in language origins. *Front. Evol. Neurosci.* 4:2. doi: 10.3389/fnevo.2012.00002

Berwick, R. C., Beckers, G. J. L., Okanoya, K., and Bolhuis, J. J. (2012). A bird's eye view of human language evolution. *Front. Evol. Neurosci.* 4:5. doi: 10.3389/fnevo.2012.00005

Fitch, W. T. (2011). The evolution of syntax: an exaptationist perspective. *Front. Evol. Neurosci.* 3:9. doi: 10.3389/fnevo.2011.00009

Friederici, A. D. (2012). Language development and the ontogeny of the dorsal pathway. *Front. Evol. Neurosci.* 4:3. doi: 10.3389/fnevo.2012.00003

Heim, S., Amunts, K., Drai, D., Eickhoff, S. B., Hautvast, S., and Grodzinsky, Y. (2012). The language–number interface in the brain: a complex parametric study of quantifiers and quantities. *Front. Evol. Neurosci.* 4:4. doi: 10.3389/fnevo.2012.00004

Kiggins, J. T., Comins, J. A., and Gentner, T. Q. (2012). Targets for a comparative neurobiology of language. *Front. Evol. Neurosci.* 4:6. doi: 10.3389/fnevo.2012.00006

Lachmann, T., Khera, G., Srinivasan, N., and van Leeuwen, C. (2012).

Learning to read aligns visual analytical skills with grapheme-phoneme mapping: evidence from illiterates. *Front. Evol. Neurosci.* 4:8. doi: 10.3389/fnevo.2012.00008

Petkov, C. I., and Jarvis, E. D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* 4:12. doi: 10.3389/fnevo.2012.00012

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Front. Evol. Neurosci.* 4:7. doi: 10.3389/fnevo.2012.00007

Rijntjes, M., Weiller, C., Bormann, T., and Musso, M. (2012). The dual loop model: its relation to language and other modalities. *Front. Evol. Neurosci.* 4:9. doi: 10.3389/fnevo.2012.00009

Rilling, J. K., Glasser, M. F., Jbabdi, S., Andersson, J., and Preuss, T. M. (2012). Continuity, divergence, and the evolution of brain language pathways. *Front. Evol. Neurosci.* 3:11. doi: 10.3389/fnevo.2011.00011

Received: 16 October 2012; accepted: 09 January 2013; published online: 28 January 2013.

Citation: Scharff C, Friederici AD and Petrides M (2013) Neurobiology of human language and its evolution: primate and non-primate perspectives. Front. Evol. Neurosci. ***5****:1. doi: 10.3389/fnevo.2013.00001*

frontiers in
EVOLUTIONARY NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 03 January 2012
doi: 10.3389/fnevo.2011.00011

Continuity, divergence, and the evolution of brain language pathways

James K. Rilling[1,2,3,4]*, Matthew F. Glasser[5], Saad Jbabdi[6], Jesper Andersson[6] and Todd M. Preuss[3,7,8]

[1] Department of Anthropology, Emory University, Atlanta, GA, USA
[2] Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, USA
[3] Center for Translational Social Neuroscience, Emory University, Atlanta, GA, USA
[4] Yerkes National Primate Research Center, Emory University, Atlanta, GA, USA
[5] Department of Anatomy and Neurobiology, Washington University, St. Louis, MO, USA
[6] Centre for Functional Magnetic Resonance Imaging of the Brain, University of Oxford, Oxford, UK
[7] Division of Neuropharmacology and Neurologic Diseases, Yerkes National Primate Research Center, Emory University, Atlanta, GA, USA
[8] Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA, USA

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

Reviewed by:
Thomas Lachmann, University of Kaiserslautern, Germany
Jonathan K. Burns, University of KwaZulu-Natal, South Africa

***Correspondence:**
James K. Rilling, Department of Anthropology, Emory University, 1557 Dickey Drive, Atlanta, GA 30322, USA.
e-mail: jrillin@emory.edu

Recently, the assumption of evolutionary continuity between humans and non-human primates has been used to bolster the hypothesis that human language is mediated especially by the ventral extreme capsule pathway that mediates auditory object recognition in macaques. Here, we argue for the importance of evolutionary divergence in understanding brain language evolution. We present new comparative data reinforcing our previous conclusion that the dorsal arcuate fasciculus pathway was more significantly modified than the ventral extreme capsule pathway in human evolution. Twenty-six adult human and twenty-six adult chimpanzees were imaged with diffusion-weighted MRI and probabilistic tractography was used to track and compare the dorsal and ventral language pathways. Based on these and other data, we argue that the arcuate fasciculus is likely to be the pathway most essential for higher-order aspects of human language such as syntax and lexical–semantics.

Keywords: language, evolution, brain, chimpanzee, arcuate fasciculus, extreme capsule

INTRODUCTION

Language is one of the fundamental evolutionary innovations of the human lineage. Our closest relatives, chimpanzees and bonobos, can learn signs, but do not produce grammatical expressions (Wallman, 1992; Rivas, 2005; Premack, 2007). How did evolution transform a non-linguistic ancestral primate brain into a linguistic human brain? The fossil record provides few clues about this transformation: we know that brain volume increased dramatically (about threefold) after the human lineage separated from that leading to chimps and bonobos, about six to eight million years ago, but soft tissues like the brain are not preserved during fossilization, so there is no record of the changes in the brain's internal organization related to language. To understand language evolution we must employ the comparative method, using information about the shared characteristics of living species to infer ancestral states (e.g., Sherwood et al., 2008; Preuss, 2011). In particular, we need to compare humans to the primates with which we are most closely related, namely apes and Old World monkeys, the latter including the familiar macaque monkeys. The scale of research done on the connections and functions of macaque brains makes them an especially valuable source of information.

NON-HUMAN PRIMATE BRAIN COMMUNICATION SYSTEMS

Intuitively, uniquely human functions would seem to require uniquely human brain structures, so some neuroscientists have maintained that the classic language areas of Broca and Wernicke must be unique to humans (e.g., Brodmann, 1909; Crick and Jones, 1993). The work of evolution, however, more commonly involves the modification of existing anatomical structures to serve different functions than the addition of new structures. There is, in fact, considerable evidence that homologs of Broca's and Wernicke's areas exist in apes and monkeys, based on similarities in architectonics, common position within the cortical mantle relative to other areas, and shared non-linguistic functions (e.g., Bonin, 1944; Galaburda and Pandya, 1982; Rizzolatti and Arbib, 1998; Preuss, 2000, 2011; Arbib, 2007). Yet, presumably, there was something about the non-human homologs of the classic language areas that made them suitable to be "recruited" (Bonin, 1944; Arbib, 2007) into the evolving language system.

Perhaps language evolved from brain systems that perform related functions in non-human primates, such as the production and perception of communicative calls and facial expressions. Area F5, the macaque homolog of the posterior part of Broca's area (area 44), is involved in the production of orofacial expressions (Petrides et al., 2005), and mirror neurons in F5 respond to communicative mouth gestures, presumably using motor simulation to form a natural link between sender and receiver that facilitates communication (Rizzolatti and Fogassi, 2007). Calls and vocalizations are processed in the ventral auditory pathway that links anterior and middle STG, STS, and inferotemporal cortex (IT) with areas 45 and 47/12 (the likely homologs of the anterior and orbital parts of Broca's area in humans) via the extreme capsule (Petrides and Pandya, 2009). This pathway is involved in auditory object identification. Although not specific for calls, both nodes (lateral belt area AL in temporal cortex and area 45 in ventrolateral PFC) include neurons that are highly responsive to species-specific vocalizations (Romanski et al., 1999). Functionally, area 45 may represent the referential meaning of calls, or may be involved in active controlled retrieval of memories associated with those calls stored in posterior cortical association areas (Petrides and Pandya, 2009). Additionally, the superior temporal gyrus appears to be left hemisphere dominant for discriminating species-specific vocalizations but not other types of auditory stimuli (Heffner and Heffner, 1986). Interestingly, in contrast to humans, in macaques the dominant prefrontal projection from posterior STG/STS is to dorsal prefrontal cortex (Petrides and Pandya, 2002), with only a minor projection to 44/45 (Deacon, 1992; Petrides and Pandya, 2009). This dorsal auditory "where" pathway carries information about the spatial location of sound (Romanski et al., 1999).

Although macaque area F5 is homologous to part of Broca's area (area 44), which plays a critical role in speech production in humans, macaque F5 does not appear to mediate production of species-specific calls, given that lesions there do not disrupt calling (Aitken, 1981). Instead, macaque calls appear to be mediated by limbic and brainstem regions and are consequently largely involuntary symptoms of specific emotional and arousal states (Deacon, 1997).

HUMAN BRAIN LANGUAGE SYSTEMS AND THEIR EVOLUTION

EVOLUTIONARY CONTINUITY

Did evolution build human language out of components of the non-human primate brain communication systems just described? If so, we would expect human language to also tap these systems. Broca's area is obviously important for human expressive communication. In addition, the ventral auditory, or extreme capsule, pathway also exists in humans (Frey et al., 2008; Makris and Pandya, 2009), extending from pars orbitalis (47) and triangularis (45) to anterior STG and then back to angular gyrus. It has been reasonably proposed that this pathway, normally involved with retrieval of memories stored in posterior association cortex, was adapted during human evolution for controlled retrieval of verbal information in the human left hemisphere (Schmahmann et al., 2007; Makris and Pandya, 2009; Petrides and Pandya, 2009). However, comparative evidence suggests that, relative to the more dorsal arcuate fasciculus pathway, this ventral pathway was not a major locus of change in human evolution.

EVOLUTIONARY DIVERGENCE

Although the human language system likely recruited components present in non-human primates, the key to understanding the evolution of human language lies not with the similarities to non-human primates but with the differences. That is, since humans possess language and other primates do not, there must be critical functional and anatomical differences between human and non-human primate brains that endow us with this special ability. We cannot determine the unique features of the human brain through human–macaque comparisons alone, as macaques are relatively distant evolutionary relatives of humans. Instead we must compare the human brain with that of our closest living relative, the chimpanzee. If we identify a characteristic in humans that is not present in chimpanzees or macaques, it is reasonable to assume that the trait uniquely evolved in humans after we diverged from chimpanzees six to eight million years ago.

HUMAN BRAIN LANGUAGE SPECIALIZATIONS

Given the traditionally accepted importance of Wernicke's and Broca's areas in language, were there changes in the temporal and frontal cortices that contain these regions? Here, we will focus on temporal cortex. Early functional MRI studies of the human visual system noted differences in the location of human and macaque visual areas (Ungerleider et al., 1998). Whereas macaque visual cortex spanned the lateral IT, human visual cortex was in a more ventral and posterior position. This prompted the suggestion that an evolutionary expansion of human language cortex in the lateral temporal lobe displaced human visual cortex to its present location. Although the visual system has not been mapped in the chimpanzee brain, the chimpanzee lunate sulcus, which marks the anterior border of V1, is in a macaque-like rather than a human-like location (Holloway et al., 2008), suggesting that chimpanzees largely preserve macaque-like visual cortical organization.

If human visual cortex was displaced by expanded temporal lobe language cortex, where specifically in the temporal lobe did this expansion take place? Lesion (Damasio et al., 1996; Dronkers et al., 2004), fMRI (Binder et al., 2009; Price, 2010), and structural and functional connectivity (Glasser and Rilling, 2008; Turken and Dronkers, 2011) data implicate the left MTG as a neural epicenter for lexical–semantic processing in the human brain (Turken and Dronkers, 2011). Functional MRI studies additionally implicate the adjacent STS as a core region involved in syntax (Grodzinsky and Friederici, 2006). If one assumes evolutionary continuity, one might reasonably hypothesize that this cortex (STS/MTG) is connected to ventrolateral prefrontal cortex via the ventral auditory pathway that was inherited from non-linguistic non-human primates. Further, this ventral pathway should mediate lexical–semantic retrieval and syntax. Given the expansion of cortical surface area (Van Essen and Dierker, 2007), we would also predict a corresponding expansion in the ventral extreme capsule pathway relative to the dorsal arcuate fasciculus pathway in linguistic humans vs. non-linguistic chimpanzees if the continuity hypothesis is correct. Furthermore, we might expect the pathway to be leftwardly asymmetric, given that lexical–semantics and syntax tend to be left-lateralized (Nucifora et al., 2005; Parker et al., 2005; Glasser and Rilling, 2008). We can test this prediction directly with comparative diffusion tractography (DT), which can estimate the extent and route of connections between cortical regions.

RESULTS AND DISCUSSION

Contrary to the hypothesis that expanded temporal lobe language cortex is most strongly connected to Broca's area via the ventral extreme capsule pathway, we previously found a qualitatively stronger connection via the dorsal arcuate fasciculus pathway (Rilling et al., 2008). These data suggest that the dorsal arcuate fasciculus pathway may have been the focus of language-related change in human evolution. To quantitatively evaluate this claim, we here compare a rough measure of connection strength of the dorsal and ventral pathways in a sample of 26 human brains with the homologous pathways in 26 chimpanzee brains. If the dorsal pathway was augmented in human evolution, then it should be stronger relative to the ventral pathway in humans vs. chimpanzees, and this is what was found. Although present in both hemispheres, the effect is more pronounced in the left hemisphere, where humans have a particularly strong dorsal pathway. Nevertheless, the dorsal pathway was leftwardly asymmetric in both species, a finding consistent with previously reported leftward asymmetries in the planum temporale, a portion of Wernicke's area (Gannon et al., 1998; Hopkins et al., 1998, 2008), and in peri-sylvian white matter volume (Cantalupo et al., 2009). These findings suggest that the anatomical substrates for lateralization of communicative functions may have been present in the common ancestor of humans and chimpanzees (Cantalupo et al., 2009). In contrast to the dorsal pathway, the ventral pathway is not asymmetric in either humans or chimpanzees. We would expect a pathway that mediates syntax and lexical–semantic retrieval to be leftwardly asymmetric, like the human arcuate, rather than symmetric, like the human extreme capsule (**Table 1**; **Figure 1**).

Finally, as reported previously (Rilling et al., 2008), in humans the arcuate projections into the temporal cortex are concentrated in STS and MTG, ventral to classic Wernicke's area, whereas in chimpanzees they are concentrated in STG. On the other hand, extreme capsule projections to temporal cortex are concentrated in STS and cortex ventral to it in both species. Thus, in terms of both pathway strength and pattern of cortical connectivity, the dorsal arcuate fasciculus seems to have undergone more evolutionary change than the ventral extreme capsule pathway.

Did the expanded arcuate fasciculus pathway displace the ventral visual stream in the human brain, as suggested above? Tracking the ventral visual stream (the inferior longitudinal fasciculus, ILF) in both species revealed that the arcuate abuts the ILF in humans but not chimps and does appear to have displaced ILF in a ventromedial direction (**Figure 2**).

CONCLUSION

Comparative DT data suggest that the specialized, derived features of human language (syntax and lexical–semantics) are likely to be mediated by the arcuate fasciculus pathway. The most cited evidence to the contrary is from a paper by Saur et al. (2010) who used fMRI to identify frontal and temporal cortical regions involved in processing word meaning and then used DT to track between these functional ROIs. They found stronger connectivity between frontal and temporal semantic ROIs via the ventral extreme capsule pathway as opposed to the dorsal arcuate fasciculus pathway. Critically, however, despite widespread activation across the MTG, they limited their tractography seeds to activation peaks in the anterior and posterior extremes of the MTG. That is, they did not track from the core lexical–semantic and syntactical areas in mid MTG and STS respectively (Vigneau et al., 2006; Glasser and Rilling, 2008; Turken and Dronkers, 2011). Furthermore, they used tensor-based single fiber tractography, which is unable to follow non-dominant pathways and gives less accurate estimates of fiber orientations (Behrens et al., 2007). Here we show

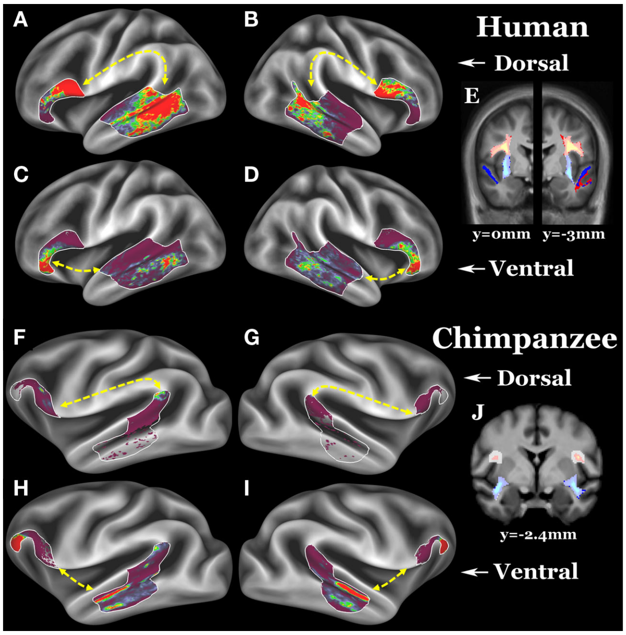


FIGURE 1 | (A–D) Group average left dorsal, right dorsal, left ventral, and right ventral pathways of 26 humans. **(E)** Left ($y = -3$ mm) and right ($y = 0$ mm) dorsal and ventral pathways in coronal slices; dorsal pathway is yellow–red, ventral pathway is light blue–blue. **(F–I)** Group average left dorsal, right dorsal, left ventral, and right ventral pathways of 26 chimpanzees. **(J)** Left and right (both $y = -2.4$ mm) dorsal and ventral pathways in coronal slices. Surface ROIs are displayed as white outlines. Fascicle selection ROIs are displayed as a translucent white layer over the pathways. For surface results, the scale is 0 (clear) to 30 (red) streamlines, for the volume results, the scale is 5 (clear) to 300 (yellow or light blue) streamlines.

Table 1 | Diffusion tractography normalized streamline counts and asymmetry indices (AIs) in chimpanzees and humans.

	Left dorsal	Right dorsal	Left ventral	Right ventral	Left D/V AI	Right D/V AI	Dorsal L/R AI	Ventral L/R AI
Human	116073	53214	27947	34753	0.61 ± 0.06**	0.00 ± 0.13	0.42 ± 0.11**	−0.17 ± 0.09
Chimpanzee	2865	379	23761	18942	−0.84 ± 0.08**	−0.88 ± 0.08**	0.66 ± 0.07**	0.08 ± 0.10
Human–Chimpanzee					1.44 ± 0.10**	0.89 ± 0.15**	−0.24 ± 0.13	−0.24 ± 0.13

Streamline counts were normalized to remove variance in ROI size (after deformation from standard ROIs to individuals) and for differences in trackability across subjects within a species. The assumption was made that the total number of streamlines counted across all four pathways should be the same across individuals within a species, as we are only interested in relative differences between the pathways across subjects and want the average normalized streamline counts to reflect equal contributions from all subjects. D, dorsal, V, ventral, L, Left, R, right, AI, Asymmetry Index [$AI_{LR} = (W_L - W_R)/(W_L + W_R)$ or $AI_{DV} = (W_D - W_V)/(W_D + W_V)$], $* = p < 0.05$, $** = p < 0.01$.

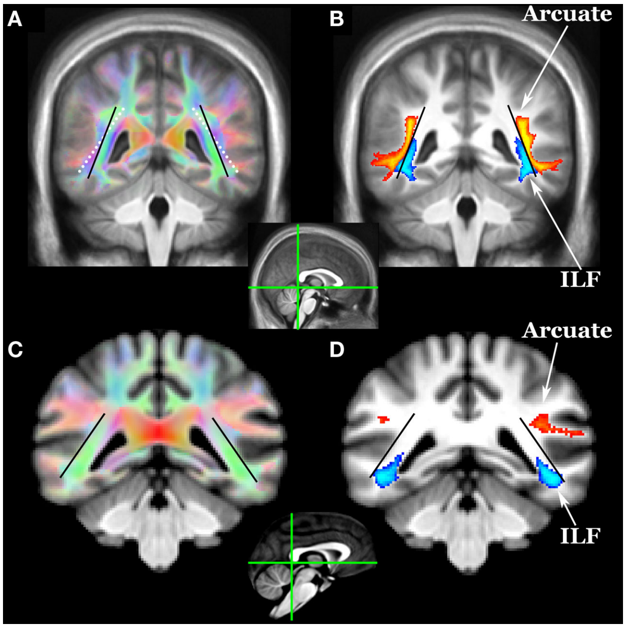


FIGURE 2 | Location of arcuate (yellow–orange) and inferior longitudinal fasciculi (ILF, blue) in (A,B) humans and (C,D) chimpanzees as revealed by diffusion tractography. Coronal sections for each species are at the posterior aspect of the splenium (see mid-sagittal insets). Tracts include voxels in which 33% or more subjects have a pathway above threshold (0.1% of waytotal). The black lines indicate the angle of the ILF in humans and chimpanzees. The white dotted line in **(A)** shows the angle of the ILF of chimps overlaid on the human color FA map. In humans, the arcuate appears to have displaced the ILF in a ventromedial direction.

that tracking from mid MTG/STS with crossing-fiber tractography yields stronger connectivity via a dorsal compared with a ventral route. This is not to say that the extreme capsule pathway has no role in human language. Indeed, there is evidence that electrical stimulation of the extreme/external capsule (EC) induces semantic paraphasias (Martino et al., 2010). Other pathways such as the inferior fronto-occipital fasciculus (IFOF) may also be involved (Duffau, 2008; Turken and Dronkers, 2011). However, we argue that the most significant modification of human brain connectivity related to language evolution, in particular the development of lexical–semantic retrieval and syntax, occurred in the arcuate fasciculus.

MATERIALS AND METHODS

SUBJECTS, ACQUISITION, AND PREPROCESSING

Twenty-six humans (17 males, mean age = 20.0, SD = 1.2) and 26 chimpanzees (26 female, mean age = 29.4, SD = 13.0) were scanned with anatomical and diffusion-weighted imaging (DWI) on Siemens 3T Trio scanners. All chimpanzee and human procedures were approved by the Emory University Animal Care and Use Committee and Institutional Review Board, respectively. Informed consent was obtained from all human subjects. The DWIs were matched across species in diffusion directions (60 $b = 1000$, 6 $b = 0$). Given their smaller brain size, chimpanzees were scanned at higher spatial resolution (1.8 vs. 2 mm isotropic for humans) and with more averages (8 vs. 2) to compensate for lower SNR. EPI distortion in chimpanzees was reduced by using a reduced FOV and matrix along the phase encoding direction to reduce the number of phase encoding steps and shorten the echo train. Following motion and eddy current correction, remaining EPI distortion was corrected using an improved version of the method of (Andersson et al., 2003). Up to three fiber orientations were estimated in each voxel using BEDPOSTx (Behrens et al., 2007). T1-weighted (T1w) anatomical images were acquired from both humans and chimpanzees with TR/TE/TI = 2600/900/3 ms and flip angle of 8°. Chimpanzees were again scanned with higher resolution (0.8 vs. 1 mm isotropic) and more averages (2 vs. 1). Additionally a single T2w average was acquired in chimpanzees at 0.8 mm resolution with otherwise identical parameters to previous human scans (Glasser and Van Essen, 2011). T1w image volume and surface processing has been previously described in humans (Glasser and Van Essen, 2011), and surfaces were made using FreeSurfer 5.1. Obtaining maximally accurate FreeSurfer surfaces in chimpanzees requires several steps outside of FreeSurfer: bias field removal, brain extraction, linear alignment to the FreeSurfer template, and changing image dimensions to 1 mm to avoid automatic resampling. The chimpanzees had a significant bias field, requiring special estimation. As the 3T T1w and T2w images have similar bias fields and inverted contrast (Glasser and Van Essen, 2011), we estimated the bias field using the approximation in Equation 1, where x and $1/x$ are the contrast for myelin in the T1w and T2w images respectively, and b is the bias field.

$$\text{sqrt}\ (\text{T1w} \times \text{T2w}) \approx \text{sqrt}\ ((x \times b) \times (1/x \times b)) = b \tag{1}$$

When restricted to brain tissue, lowpass filtering "b" produces an accurate bias field estimate. A non-linear volumetric chimpanzee template was previously generated (Li et al., 2010) and we iteratively generated a chimpanzee surface template with standard energy-based FreeSurfer registration. Chimpanzee myelin maps were generated using methods described previously in humans (Glasser and Van Essen, 2011) and human myelin maps were from that study.

TRACTOGRAPHY METHODS

Our goal was to track between Broca's region (i.e., area 44, 45, and 47l) and association cortex in the posterior two-thirds of the lateral temporal cortex lying dorsal and anterior to visual association cortex and ventral to early auditory cortex. Frontal and temporal surface ROIs (white outlines in **Figures 1A–D, G–I**) were used together with volumetric fascicle selection ROIs (translucent white on coronal slice in **Figures 1E,J**) that required streamlines to travel via either a dorsal or ventral route. ROIs were drawn on group average templates and then warped into individual subjects' diffusion space for tractography.

Surface ROIs were defined as follows: Fiber pathways of interest were initially localized by tracking from white matter ROIs in the superior longitudinal fasciculus (SLF) and EC. The surface terminations from this tractography defined an outer bound on the possible connections between frontal and temporal regions, and, within this area, myelin maps and probabilistic cytoarchitecture were used to define homologous frontal and temporal surface ROIs across hemispheres and species. The frontal surface ROI was defined in humans using surface-based probabilistic cytoarchitectonic areas 44, 45, and 47l (Amunts et al., 1999; Öngür et al., 2003; Fischl et al., 2008; Van Essen et al., 2011) and was located in a region of lightly myelinated cortex posterior/superior to heavily myelinated area 47 m on the inferior frontal gyrus. In chimpanzees, volume-based probabilistic areas 44 and 45 (Schenker et al., 2010) together with cortical myelination were used to define a homologous region. Single ROI tractography from this region was used in both species to further refine the localization of temporal terminations. The lightly myelinated posterior two-thirds of the lateral temporal cortex in the STG, STS, and MTG including probabilistic areas TE 3.0 (Morosan et al., 2005) in humans and 22 (Spocter et al., 2010) in chimps that was bordered superiorly by more myelinated auditory belt cortex, posteriorly by more myelinated MT+ cortex, and ventrally by more myelinated ventral visual cortex formed the temporal surface ROI. These ROIs were constrained to include only those vertices that also received surface terminations in the localizer tractography.

The resulting surface ROIs were largely the same shape and size across hemispheres, but differed across species. As has been previously suspected for macaque monkeys (Ungerleider et al., 1998; Van Essen and Dierker, 2007), temporal cortical areas have undergone significant shifts relative to the cortical geography in humans relative to chimpanzees (Glasser et al., 2011), and geographically corresponding ROIs (i.e., ROIs of the same shape and size) would not have spanned homologous cortex. The availability of human and chimpanzee surface templates with rich probabilistic post-mortem and *in vivo* architectonic data is unprecedented for a non-invasive connectivity study.

The final probabilistic tractography was constrained to run symmetrically via either the dorsal or ventral route between the surface ROIs and streamlines were displayed on the surface (terminations) and in the volume (fascicles). 150,000 streamlines were sent out from each vertex/voxel in proportion to the fiber volume fraction in voxels with more than one fiber modeled and streamlines were stopped when they attempted to exit the white matter surface. The total number of streamlines that successfully traced the required route (the "waytotal") was recorded during tractography. Within a subject, these waytotals are proportional to the probability that the streamlines reach their target ROIs, and provide a rough metric of pathway strength when compared to another pathway seeded from ROIs of the same size. To compare across individuals, however, it is necessary to normalize these waytotals by the size of the ROIs used as seeds and the total number of streamlines counted across all four pathways. This normalization accounts for differences in ROI size after deforming standard ROIs to individuals and for global differences in trackability between individuals (e.g., motion, SNR, brain size) within a species. AIs were used (see **Table 1** for values and definitions), and the surface terminations and volume probabilistic fascicles were also normalized by the sum of each subject's waytotals so each contributed equally to the group average (**Figure 1**). A one-sample t-test (two tailed) was used to test if each AI was significantly different from zero (no asymmetry), and a two-sample t-test (two tailed) was used to test if the AIs were different between humans and chimpanzees.

The ILF (**Figure 2**) was defined using two volume ROIs orthogonal to the pathway one-third of the way back from the temporal pole and in the deep occipital white matter. The atlas brain was rotated 45° around the x-axis so that a coronal section cut the ILF orthogonally in the anterior temporal lobe. An ROI was drawn within the entire white matter on this slice, and single ROI tractography was done to identify occipital projections. A second ROI was drawn to select these projections, and the result in **Figure 2** was produced with symmetric two ROI tractography between these ROIs.

ACKNOWLEDGMENTS

We thank Ashley DeMarco, Longchuan Li, Govind Bhagavatheeshwaran, Bhargav Errangi, Xiaodong Zhang, and Xiaoping Hu for assistance with various aspects of this study. Matthew F. Glasser was supported by a National Research Science Award – Medical Scientist NIH T32 GM007200. Mark Jenkinson provided a preview version of boundary-based registration in FLIRT for T2w to T1w registration. Bill Hopkins provided probabilistic cytoarchitectonic areas for the chimpanzees. Some computations were performed using facilities of the Washington University Center for High Performance Computing, partially supported by Grant NCRR 1S10RR022984-01. Funding was provided by NIMH Grant R01 MH084068-01A1, NIA Grant 5P01 AG026423-03, and the Yerkes Base Grant: NIH RR-00165. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

REFERENCES

Aitken, P. G. (1981). Cortical control of conditioned and spontaneous vocal behavior in rhesus monkeys. *Brain Lang.* 13, 171–184.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H., and Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412, 319–341.

Andersson, J. L. R., Skare, S., and Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* 20, 870–888.

Arbib, M. A. (2007). "Premotor cortex and the mirror neuron hypothesis for the evolution of language," in *Evolution of Nervous Systems, Vol. 4, Primates*, eds J. H. Kaas and T. M. Preuss (Oxford: Elsevier), 417–422.

Behrens, T., Berg, H. J., Jbabdi, S., Rushworth, M., and Woolrich, M. (2007). Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* 34, 144–155.

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796.

Bonin, G. V. (1944). "The architecture," in *The Precentral Motor Cortex*, ed. P. C. Bucy (Urbana, IL: University of Illinois Press), 7–82.

Brodmann, K. (1909). Vergleichende Lokalisationslehre der Grosshirnrhinde. Leipzig: Barth (reprinted as Brodmann's "Localisation in the Cerebral Cortex," translated and edited by L. J. Garey, London: Smith-Gordon, 1994).

Cantalupo, C., Oliver, J., Smith, J., Nir, T., Taglialatela, J. P., and Hopkins, W. D. (2009). The chimpanzee brain shows human-like perisylvian asymmetries in white matter. *Eur. J. Neurosci.* 30, 431–438.

Crick, F., and Jones, E. (1993). Backwardness of human neuroanatomy. *Nature* 361, 109–110.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., and Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature* 380, 499–505.

Deacon, T. (1997). *The Symbolic Species.* New York: W. W. Norton.

Deacon, T. W. (1992). Cortical connections of the inferior arcuate sulcus cortex in the macaque brain. *Brain Res.* 573, 8–26.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D. Jr., Redfern, B. B., and Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition* 92, 145–177.

Duffau, H. (2008). The anatomo-functional connectivity of language revisited. New insights provided by electrostimulation and tractography. *Neuropsychologia* 46, 927–934.

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B., Mohlberg, H., Amunts, K., and Zilles, K. (2008). Cortical folding patterns and predicting cytoarchitecture. *Cereb. Cortex* 18, 1973.

Frey, S., Campbell, J. S., Pike, G. B., and Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.* 28, 11435–11444.

Galaburda, A. M., and Pandya, P. N. (1982). "Role of architectonics and connections in the study of primate brain evolution," in *Primate Brain Evolution: Methods and Concepts*, eds E. Armstrong and D. Falk (New York: Plenum), 203–226.

Gannon, P. J., Holloway, R. L., Broadfield, D. C., and Braun, A. R. (1998). Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. *Science* 279, 220–222.

Glasser, M., Preuss, T., Snyder, L., Nair, G., Rilling, J., Zhang, X., Li, L., and Van Essen, D. (2011). *Comparative Mapping of Cortical Myelin Content in Humans, Chimpanzees, and Macaques Using T1-weighted and T2-weighted MRI.* Washington, DC: Society for Neuroscience.

Glasser, M. F., and Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cereb. Cortex* 18, 2471–2482.

Glasser, M. F., and Van Essen, D. C. (2011). Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* 31, 11597–11616.

Grodzinsky, Y., and Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. Curr Opin Neurobiol 16, 240–246.

Heffner, H. E., and Heffner, R. S. (1986). Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by Japanese macaques. *J. Neurophysiol.* 56, 683–701.

Holloway, R. L., Sherwood, C. C., Rilling, J. K., and Hof, P. R. (2008). "Evolution, of the brain: in humans – paleoneurology," in *Encyclopedia of Neuroscience*, eds M. D. Binder, N. Hirokawa, U. Windhorst, and M. C. Hirsch (Springer-Verlag), 1326–1334.

Hopkins, W. D., Marino, L., Rilling, J. K., and Macgregor, L. A. (1998). Planum temporale asymmetries in great apes as revealed by magnetic resonance imaging (MRI). *Neuroreport* 9, 2913–2918.

Hopkins, W. D., Taglialatela, J. P., Meguerditchian, A., Nir, T., Schenker, N. M., and Sherwood, C. C. (2008). Gray matter asymmetries in chimpanzees as revealed by voxel-based morphometry. *Neuroimage* 42, 491–497.

Li, L., Preuss, T. M., Rilling, J. K., Hopkins, W. D., Glasser, M. F., Kumar, B., Nana, R., Zhang, X., and Hu, X. (2010). Chimpanzee (*Pan troglodytes*) precentral corticospinal system asymmetry and handedness: a diffusion magnetic resonance imaging study. *PloS ONE* 5, e12886. doi:10.1371/journal.pone.0012886

Makris, N., and Pandya, D. N. (2009). The extreme capsule in humans and rethinking of the language circuitry. *Brain Struct. Funct.* 213, 343–358.

Martino, J., Brogna, C., Robles, S. G., Vergani, F., and Duffau, H. (2010). Anatomic dissection of the inferior fronto-occipital fasciculus revisited in the lights of brain stimulation data. *Cortex* 46, 691–699.

Morosan, P., Schleicher, A., Amunts, K., and Zilles, K. (2005). Multimodal architectonic mapping of human superior temporal gyrus. *Anat. Embryol. (Berl).* 210, 401–406.

Nucifora, P. G., Verma, R., Melhem, E. R., Gur, R. E., and Gur, R. C. (2005). Leftward asymmetry in relative fiber density of the arcuate fasciculus. *Neuroreport* 16, 791–794.

Öngür, D., Ferry, A., and Price, J. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex. *J. Comp. Neurol.* 460, 425–449.

Parker, G. J., Luzzi, S., Alexander, D. C., Wheeler-Kingshott, C. A., Ciccarelli, O., and Lambon Ralph, M. A. (2005). Lateralization of ventral and dorsal auditory-language pathways in the human brain. *Neuroimage* 24, 656–666.

Petrides, M., Cadoret, G., and Mackey, S. (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature* 435, 1235–1238.

Petrides, M., and Pandya, D. N. (2002). "Association pathways of the prefrontal cortex and functional observations," in *Principles of Frontal Lobe Function*, eds D. T. Stuss and R. T. Knight (New York: Oxford University Press), 31–50.

Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7, e1000170. doi:10.1371/journal.pbio.1000170

Premack, D. (2007). Human and animal cognition: continuity and discontinuity. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13861–13867.

Preuss, T. M. (2000). "What's human about the human brain?" in *The New Cognitive Neurosciences*, 2nd Edn., ed. Michael S. Gazzaniga (Cambridge, MA: MIT Press), 1219–1234.

Preuss, T. M. (2011). The human brain: rewired and running hot. *Ann. N. Y. Acad. Sci.* 1225(Suppl. 1), E182–E191.

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. N. Y. Acad. Sci.* 1191, 62–88.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., and Behrens, T. E. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.

Rivas, E. (2005). Recent use of signs by chimpanzees (*Pan troglodytes*) in interactions with humans. *J. Comp. Psychol.* 119, 404–417.

Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194.

Rizzolatti, G., and Fogassi, L. (2007). "Mirror neurons and social cognition," in *The Oxford Handbook of Evolutionary Psychology*, eds R. I. M. Dunbar and L. Barrett (Oxford: Oxford University Press), 179–196.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136.

Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Kupper, H., Kellmeyer, P., Kummerer, D., Kloppel, S., Glauche, V., Lange, R., Mader, W., Feess, D., Timmer, J., and Weiller, C. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *Neuroimage* 49, 3187–3197.

Schenker, N. M., Hopkins, W. D., Spocter, M. A., Garrison, A. R., Stimpson, C. D., Erwin, J. M., Hof, P. R., and Sherwood, C. C. (2010). Broca's area homologue in chimpanzees (*Pan troglodytes*): probabilistic mapping, asymmetry, and comparison to humans. *Cereb. Cortex* 20, 730.

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., De Crespigny, A. J., and Wedeen, V. J. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* 130, 630–653.

Sherwood, C. C., Subiaul, F., and Zawidzki, T. W. (2008). A natural history of the human mind: tracing evolutionary changes in brain and cognition. *J. Anat.* 212, 426–454.

Spocter, M. A., Hopkins, W. D., Garrison, A. R., Bauernfeind, A. L., Stimpson, C. D., Hof, P. R., and Sherwood, C. C. (2010). Wernicke's area homologue in chimpanzees (*Pan troglodytes*) and its relation to the appearance of modern human language. *Proc. R. Soc. Lond. B Biol. Sci.* 277, 2165.

Turken, A. U., and Dronkers, N. F. (2011). The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. *Front. Syst. Neurosci.* 5:1. doi:10.3389/fnsys.2011.00001

Ungerleider, L. G., Courtney, S. M., and Haxby, J. V. (1998). A neural system for human visual working memory. *Proc. Natl. Acad. Sci. U.S.A.* 95, 883–890.

Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J., and Coalson, T. (2011). Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cereb. Cortex.* doi: 10.1093/cercor/bhr291

Van Essen, D. C., and Dierker, D. L. (2007). Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron* 56, 209–225.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., and Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30, 1414–1432.

Wallman, J. (1992). *Aping Language.* Cambridge: Cambridge University Press.

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: 31 October 2011; paper pending published: 26 November 2011; accepted: 12 December 2011; published online: 03 January 2012.

*Citation: Rilling JK, Glasser MF, Jbabdi S, Andersson J and Preuss TM (2012) Continuity, divergence, and the evolution of brain language pathways. Front. Evol. Neurosci. **3**:11. doi: 10.3389/fnevo.2011.00011*

frontiers in
EVOLUTIONARY NEUROSCIENCE

PERSPECTIVE ARTICLE
published: 15 May 2012
doi: 10.3389/fnevo.2012.00007

Ventral and dorsal streams in the evolution of speech and language

Josef P. Rauschecker[1,2]*

[1] Laboratory of Integrative Neuroscience and Cognition, Department of Neuroscience, Georgetown University Medical Center, Washington, DC, USA
[2] Mind/Brain Laboratory, Department of Biomedical Engineering and Computational Science, Aalto University, Helsinki, Finland

Edited by:
Angela D. Friederici, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Angela D. Friederici, Max Planck Institute for Human Cognitive and Brain Sciences, Germany
Richard J. S. Wise, Imperial College London, UK

***Correspondence:**
Josef P. Rauschecker, Laboratory of Integrative Neuroscience and Cognition, Department of Neuroscience, Georgetown University Medical Center, 3970 Reservoir Road, N.W., Washington, DC 20057-1460, USA.
e-mail: rauschej@georgetown.edu

The brains of humans and old-world monkeys show a great deal of anatomical similarity. The auditory cortical system, for instance, is organized into a ventral and a dorsal pathway in both species. A fundamental question with regard to the evolution of speech and language (as well as music) is whether human and monkey brains show principal differences in their organization (e.g., new pathways appearing as a result of a single mutation), or whether species differences are of a more subtle, quantitative nature. There is little doubt about a similar role of the ventral auditory pathway in both humans and monkeys in the decoding of spectrally complex sounds, which some authors have referred to as auditory object recognition. This includes the decoding of speech sounds ("speech perception") and their ultimate linking to meaning in humans. The originally presumed role of the auditory dorsal pathway in spatial processing, by analogy to the visual dorsal pathway, has recently been conceptualized into a more general role in sensorimotor integration and control. Specifically for speech, the dorsal processing stream plays a role in speech production as well as categorization of phonemes during on-line processing of speech.

Keywords: cerebral cortex, macaque monkey, human, communication sounds, speech, music, internal models, brain connectivity

From an auditory point of view, spoken language starts with the processing of complex auditory signals. Physiological recordings in non-human primates suggest that neurons already at the secondary stage of processing along the auditory cortical pathway (the lateral belt areas) can show a preference for species-specific communication calls (Rauschecker et al., 1995). This response tuning is generated by convergence of input from lower-order neurons that respond to simple sounds like tones, frequency-modulated sweeps, or band-passed noise bursts. Neurons are sensitive to highly specific combinations of such inputs, and combining signals in a non-linear conjunctive AND-logic leads to the existence of neurons that respond specifically to certain types of calls. There is no reason to believe that the human auditory cortex does not contain similar neurons with combination sensitivity and a similar hierarchy from rather simple to more complex neurons, whose incidence increases from primary auditory cortex to more anterior regions of the superior temporal lobe (Rauschecker, 1998; Rauschecker and Tian, 2000).

Indeed, early studies of human auditory cortex with functional magnetic resonance imaging (fMRI) have shown that primary auditory cortex responds best to tones, while at the next stage, the equivalent of the lateral belt in the monkey, band-passed noise bursts are more effective stimuli (Wessinger et al., 2001). Further along the antero-ventral pathway, cortical regions are selectively activated by words and intelligible speech sounds (Binder et al., 2000; Scott et al., 2000). This hierarchical organization of the auditory ventral stream with regard to speech-sound processing was recently corroborated with more refined techniques (Chevillet et al., 2011b). Furthermore, a meta-analysis of more than 100 neuroimaging studies of human speech processing has demonstrated that cortical regions in the mid-STG near the human lateral belt are sensitive to phonemes; farther afield in anterior STG, words are processed; finally, in the most anterior locations of STS, short phrases lead to selective activation (DeWitt and Rauschecker, 2012).

Invariant representation of sounds is another important step toward establishing a usable system for auditory communication, such as speech. There is evidence that invariances are formed along the antero-ventral stream as well (DeWitt and Rauschecker, 2012). However, other reports have found that premotor regions may be involved too (e.g., Chevillet et al., 2011a; Lee et al., 2012). It appears possible, therefore, that invariances are formed in different ways: once on the basis of spectro-temporal information, which is pooled along the frequency domain in the sense of an OR-logic within the auditory ventral stream; and independently in the domain of motor gestures, which are formed originally for speech production, but are invoked during the processing of speech as well. The same is almost inevitably true for the processing of other complex sounds that can be classified into discrete categories (Leaver and Rauschecker, 2010). Such auditory objects are also represented in anterior regions of the STG, but premotor cortex participates in their encoding as long as they can be produced and thus invoke a motor code. Monkeys are naturally handicapped by their less sophisticated vocal apparatus, which limits their vocal repertoire and their capacity to mimic sounds. The involvement of the dorsal pathway (including premotor regions) in the processing and categorization of self-produced sounds will, therefore, have to be tested by other means (Remedios et al., 2009).

The involvement of the dorsal auditory pathway, including premotor and inferior parietal regions, in the encoding and representation of temporally extended sounds (or sound sequences) became especially evident, when imagery of musical melodies was investigated (Leaver et al., 2009). During the learning of such sequences, the basal ganglia were actively engaged, whereas after these sequences became highly familiar, the same sequences activated more and more prefrontal areas. It appears, therefore, that the basal ganglia are responsible for the concatenation of sequential auditory information or formation of "chunks," which represent information about conditional probabilities for one sound being followed by another. Once the chunks have been formed, they are once again stored in prefrontal regions. A similar chunking process occurs with cued sequences of learned finger movements (Koechlin and Jubault, 2006). This process involves prefrontal cortex near Broca's area and has, therefore, been compared with models of language (Hagoort, 2005), redefining Broca's area in terms of chunking ("unification") of semantic, syntactic, and phonological information.

Thus, the role of the dorsal stream can be conceptualized into one of sensorimotor integration and control and applies to all kinds of sequential stimuli, even beyond the auditory domain. Specifically for speech, the dorsal processing stream plays a role in speech production as well as categorization of phonemes during on-line processing of speech (Rauschecker and Scott, 2009; Rauschecker, 2011; **Figure 1**). The former role conforms to the classical idea of an "efference copy" or feed-forward model and allows for fast and efficient on-line control of speech production. By contrast, the latter function can be formalized as an inverse model during real-time speech processing, creating the affordances of the speech signal in a Gibsonian sense (Gibson, 1966; Rauschecker, 2005). Both functions require a (direct or indirect) connection between sensory and motor cortical structures of the brain, whereby subcortical structures (e.g., the basal ganglia) provide an additional link setting up transitional probabilities during associative learning of sound sequences.

Comparing human and monkey brain connectivity along the dorsal stream, there may be quantitative differences in the strengths of these connections, but there does not seem to be a difference in principle (Frey et al., 2008). Similarly, in the ventral stream, the fine-grain organization of cortical areas and the fine-tuning of its neuronal elements may be richer in humans than in monkeys, providing humans with a perceptual network for the detection of more subtle differences in the acoustic signal. The decisive distinction between humans and monkeys may,

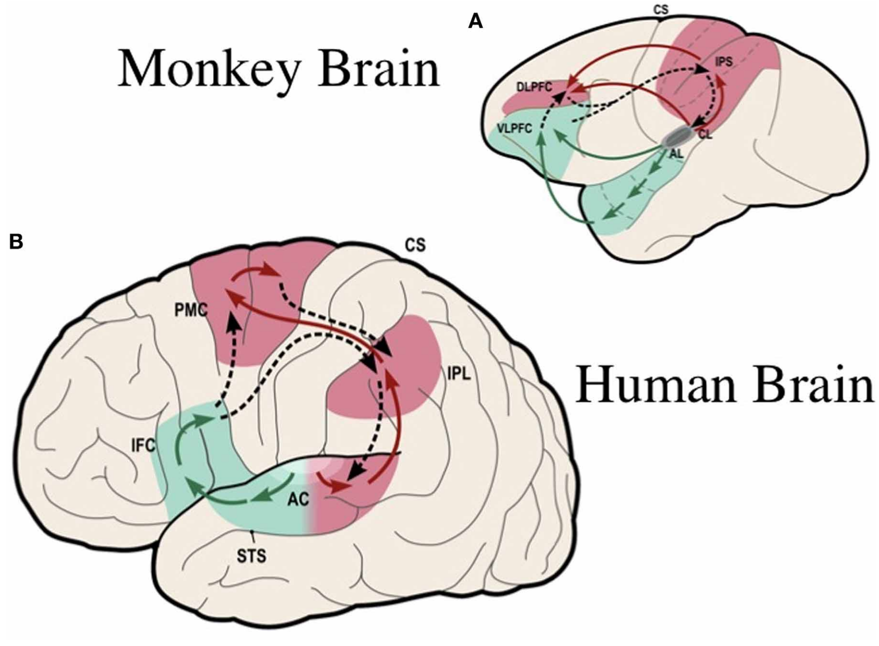


FIGURE 1 | Ventral and dorsal streams for the processing of complex sounds in the primate brain: (A) in the rhesus monkey [modified from Rauschecker and Tian (2000)]; **(B)** in the human [simplified from Rauschecker and Scott (2009)]. The ventral stream (in green) plays a general role in auditory object recognition, including perception of vocalizations and speech. The dorsal stream (in red) pivots around inferior/posterior parietal cortex, where a quick sketch of sensory event information is compared with an efference copy of motor plans (dashed lines). Thus, the dorsal stream plays a general role in sensorimotor integration and control. In clockwise fashion, starting out from auditory cortex, the processing loop performs as a forward model: object information, such as vocalizations and speech, is decoded in the antero-ventral stream all the way to category-invariant inferior frontal cortex (IFC, or VLPFC in monkeys) and transformed into articulatory representations (DLPFC or ventral PMC). Frontal activations are transmitted to the IPL and pST, where they are compared with auditory and other sensory information. AC, auditory cortex; AL, antero-lateral area; CL, caudo-lateral area; STS, superior temporal sulcus; IFC, inferior frontal cortex; DLPFC, VLPFC, dorsolateral and ventrolateral prefrontal cortex; PMC, premotor cortex; IPL, inferior parietal lobule; IPS, inferior parietal sulcus; CS, central sulcus; pST, posterior superior temporal region. [Composite figure adapted, with permission, from Rauschecker (2011)] however, lie in a third component where ventral and dorsal streams converge and interact: the prefrontal network. With its own hierarchical organization it provides the substrate for recursive processing of nested sequences, as they are typical for human grammatical language structures (Friederici, 2004). Again, however, this emergent new ability of humans may be based on a quantitative rather than principal difference in human and monkey brain organization, which ties in the existing strengths of both ventral and dorsal processing streams with fronto-parietal networks underlying working memory.

To test the real evolutionary similarity of human and monkey ventral and dorsal streams, two things have to happen in future studies:

(1) Connectivity studies in both species have to investigate in great detail which areas are connected. This will establish a greater amount of homology than other approaches, especially when the same techniques of structural and functional imaging are utilized. While anatomical tracer studies in monkeys will remain the gold standard (Romanski et al., 1999; Petrides and Pandya, 2009; Hackett, 2011), non-invasive fiber tractography using MRI-based technology will gain increasing importance as its resolution improves, because the exact same approach can be used in both species. Early attempts using diffusion tensor imaging (DTI) have had insufficient power to resolve crossing fibers within a single voxel or disentangle fibers with crossing trajectories (Catani et al., 2005; Croxson et al., 2005; Anwander et al., 2007; Rilling et al., 2008). Such studies have, therefore, remained inconclusive with regard to monkey-human homologies in language evolution. High-angular-resolution techniques, such as diffusion spectrum imaging (DSI), have been utilized successfully in humans (e.g., Frey et al., 2008) and in monkeys (Schmahmann et al., 2007; Wedeen et al., 2008). Cross-validation studies of autoradiographic tract tracing and DSI in monkeys have shown a remarkable concordance of results between tracer studies and DSI (Schmahmann et al., 2007). However further improvements in resolution and reductions in scan time are certainly needed and possible, before DSI studies can become routine. Functional studies based on blood-oxygenation-level-dependent (BOLD) responses are feasible in both species as well (Petkov et al., 2006) and can elucidate connectivity to a certain extent. Microstimulation techniques as another approach to analyze connectivity (Kikuchi et al., 2008), on the other hand, are limited to animal studies.

(2) Behavioral monkey studies have to be designed that test the above concepts and go beyond traditional models. "What" and "where" processing are still characteristic for the two streams, but as generalized models are developed (Rauschecker and Scott, 2009; Rauschecker, 2011), more appropriate monkey studies have to follow. These studies have to focus on the computational transformations that occur between the various processing stages rather than merely the connectivity describing different anatomical pathways.

ACKNOWLEDGMENTS

The collection of the underlying data as well as the writing of this article were supported by grants from the National Science Foundation (BCS-0519127, OISE-0730255), from the National Institutes of Health (R01NS052494, RC1DC010720), and from the Academy of Finland (FiDiPro).

REFERENCES

Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., and Knösche, T. R. (2007). Connectivity-based parcellation of Broca's area. *Cereb. Cortex* 17, 816–825.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528.

Catani, M., Jones, D. K., and ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.

Chevillet, M. A., Jiang, X., Rauschecker, J. P., and Riesenhuber, M. (2011a). Automatic phoneme categorization in the dorsal auditory pathway. *Soc. Neurosci. Abstr.* 172.09.

Chevillet, M., Riesenhuber, M., and Rauschecker, J. P. (2011b). Functional localization of the ventral auditory "what" stream hierarchy. *J. Neurosci.* 31, 9345–9352.

Croxson, P. L., Johansen-Berg, H., Behrens, T. E. J., Robson, M. D., Pinsk, M. A., Gross, C. G., Richter, W., Richter, M. C., Kastner, S., and Rushworth, M. F. S. (2005). Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. *J. Neurosci.* 25, 8854–8866.

DeWitt, I., and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* 109, E505–E514.

Frey, S., Campbell, J. S., Pike, G. B., and Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.* 28, 11435–11444.

Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends Cogn. Sci.* 8, 245–247.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems.* London: Allen and Unwin.

Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hear. Res.* 271, 133–146.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423.

Kikuchi, Y., Rauschecker, J. P., Mishkin, M., Augath, M., Logothetis, N. K., and Petkov, C. I. (2008). Voice region connectivity in the monkey assessed with microstimulation and functional imaging. *Soc. Neurosci.* 34, 850.2.

Koechlin, E., and Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron* 50, 963–974.

Leaver, A., van Lare, J. E., Zielinski, B. A., Halpern, A., and Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *J. Neurosci.* 29, 2477–2485.

Leaver, A. M., and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612.

Lee, Y.-S., Turkeltaub, P., Granger, R., and Raizada, R. D. S. (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *J. Neurosci.* 32, 3942–3948.

Petkov, C. I., Kayser, C., Augath, M., and Logothetis, N. K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol.* 4:e215. doi: 10.1371/journal.pbio.0040215

Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7:e170. doi: 10.1371/journal.pbio.1000170

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Curr. Opin. Neurobiol.* 8, 516–521.

Rauschecker, J. P. (2005). Vocal gestures and auditory objects. *Behav. Brain Sci.* 28, 143–144.

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor integration and control. *Hear. Res.* 271, 16–25.

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: non-human primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.

Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11800–11806.

Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114.

Remedios, R., Logothetis, N. K., and Kayser, C. (2009). Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18010–18015.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., and Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136.

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., de Crespigny, A. J., and Wedeen, V. J. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* 130, 630–653.

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–293.

Wedeen, V. J., Wang, R. P., Schmahmann, J. D., Benner, T., Tseng, W. Y., Dai, G., Pandya, D. N., Hagmann, P., D'Arceuil, H., and de Crespigny, A. J. (2008). Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *Neuroimage* 41, 1267–1277.

Wessinger, C. M., van Meter, J., Tian, B., van Lare, J., Pekar, J., and Rauschecker, J. P. (2001). Hierarchical organization of human auditory cortex revealed by functional magnetic resonance imaging. *J. Cogn. Neurosci.* 13, 1–7.

Conflict of Interest Statement: The author declares 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: 01 November 2011; paper pending published: 15 December 2011; accepted: 25 April 2012; published online: 15 May 2012.

Citation: Rauschecker JP (2012) Ventral and dorsal streams in the evolution of speech and language. Front. Evol. Neurosci. ***4****:7. doi: 10.3389/fnevo. 2012.00007*

frontiers in
EVOLUTIONARY NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 06 February 2012
doi: 10.3389/fnevo.2012.00003

Language development and the ontogeny of the dorsal pathway

Angela D. Friederici *

Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Edited by:
Constance Scharff, Freie Universitaet Berlin, Germany

Reviewed by:
Steven Chance, Oxford University, UK
Paul M. Nealen, Indiana University of Pennsylvania, USA

***Correspondence:**
Angela D. Friederici, Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1A, 04103 Leipzig, Germany.
e-mail: angelafr@cbs.mpg.de

In the absence of clear phylogenetic data on the neurobiological basis of the evolution of language, comparative studies across species and across ontogenetic stages within humans may inform us about the possible neural prerequisites of language. In the adult human brain, language-relevant regions located in the frontal and temporal cortex are connected via different fiber tracts: ventral and dorsal pathways. Ontogenetically, it has been shown that newborns display an adult-like ventral pathway at birth. The dorsal pathway, however, seems to display two subparts which mature at different rates: one part, connecting the temporal cortex to the premotor cortex, is present at birth, whereas the other part, connecting the temporal cortex to Broca's area, develops much later and is still not fully matured at the age of seven. At this age, typically developing children still have problems in processing syntactically complex sentences. We therefore suggest that the mastery of complex syntax, which is at the core of human language, crucially depends on the full maturation of the fiber connection between the temporal cortex and Broca's area.

Keywords: grammar, development, fiber tract, arcuate fasciculus

The neural basis of language evolution must remain speculative, since clear phylogenetic data are unavailable. However, there are two alternative, though more indirect ways, to approach this issue. One approach is to compare different species in their ability to learn language, in particular, syntax or rule-based sequences. A second is to consult ontogenetic data on language development and brain maturation, under the assumption that ontogeny to some extent reflects phylogeny. In this article, data from both approaches, with a strong focus on rule-based and syntactic sequence learning, will be discussed.

Central to the discussion is not only whether such sequences can be learned, but more crucially, what type of syntactic sequence can be learned. A fundamental distinction has been made between two grammar types, namely finite state grammars (FSG) following an $(AB)^n$ rule and phrase structure grammar (PSG) following an A^nB^n rule (Hauser et al., 2002; Fitch and Hauser, 2004; see **Figure 1**).

There are at least three possible mechanisms through which grammatical sequence learning can take place: (1) adjacent dependencies, as in $(AB)^n$ grammars, and also non-adjacent dependencies, which do not involve higher-order hierarchies, could be learned by extracting phonological regularities from the auditory input and memorizing these for further use; (2) adjacent dependencies between A and B in $(AB)^n$ grammars or between a determiner, e.g., *the* and a noun, e.g., *man* in natural grammars could be learned through the same mechanism described in (1), but without the buildup of a minimal hierarchy or (3) through the computation "Merge" that binds two elements into a minimal hierarchical structure (Chomsky, 1995) the basic mechanism to learn a natural grammar with its asymmetric higher-order hierarchical structure. This requires a computational system that goes beyond the mechanisms described in (1) and (2), and requires the computation Merge more than once.

In the following section, we will review the success of grammar learning in different species, discuss the possible underlying processing mechanisms, and debate their neural basis. The data from these studies examining this suggest that the three grammar learning mechanisms described above can be related to three different neural circuits: (1) an input-to-output circuit present in vocal learning animals, (2) a circuit subserving the learning of $(AB)^n$ structures, and (3) a circuit involving the learning of A^nB^n structures.

GRAMMAR LEARNING ACROSS SPECIES

There are several studies that have taken a species-comparative approach. Some have compared artificial grammar learning between human and non-human primates, or have used similar grammar types to investigate songbirds' ability to learn grammatical sequences. Others have additionally discussed the neural basis of these learning abilities.

Fitch and Hauser (2004) were the first to investigate grammar learning in human and non-human primates using FSG and PSG grammars (**Figure 1**). Testing cotton-top tamarins and human adults in a behavioral grammar learning study, they found that humans could learn both grammar types easily, whereas the monkeys were only able to learn the FSG. The neural basis for this ability in cotton-top tamarins is unknown, since there are no functional or structural brain studies on this type of monkey. There are, however, a number of structural imaging studies on macaques, chimpanzees, and humans (Catani et al., 2002; Anwander et al., 2007; Rilling et al., 2008; Saur et al., 2008; Makris and Pandya, 2009; Petrides and Pandya, 2009). These studies indicate that the frontal and temporal regions which are known to be involved

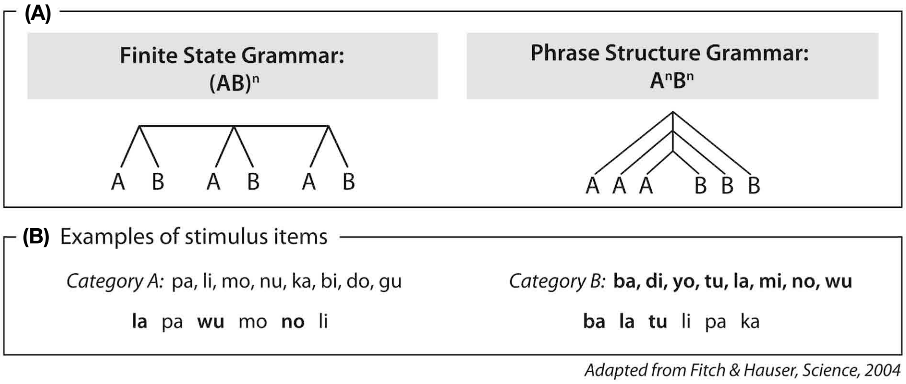


FIGURE 1 | Artificial grammar used in Fitch and Hauser (2004). (A) Structure of sequences. **(B)** Category A syllables and Category B syllables used in the sequences as well as examples of an $(AB)^n$ sequence (left panel) and an A^nB^n sequence (right panel). Category A syllables were produced by a female speaker, Category B syllables by a male speaker. Category membership was thus coded by the pitch of voice.

in language processing in humans are connected via ventral and dorsal fiber bundles in both humans and non-human primates. A direct comparison, however, revealed differences between humans and non-human primates; macaques and chimpanzees display a strong ventral and a weak dorsal pathway, whereas humans display a strong dorsal pathway and a well-developed ventral pathway. The dorsal pathway was, therefore, discussed as the crucial pathway for the language ability in humans (Rilling et al., 2008).

This difference in the structure of these pathways between humans and non-human primates is of particular interest in the light of a functional and structural imaging study in humans (Friederici et al., 2006), which applied the same artificial grammar types as used in the behavioral study by Fitch and Hauser (2004). In humans, the $(AB)^n$ grammar, with its adjacent dependencies, activated the frontal operculum, which is connected via the ventral pathway to the temporal cortex. Interestingly, the A^nB^n grammar additionally recruited Broca's area, which is connected to the temporal cortex via the dorsal pathway (Friederici et al., 2006). These data were taken to suggest that Broca's area and its dorsal connection to the temporal cortex, in particular, supports the processing of higher-order hierarchically structured sequences relevant to language.

This conclusion, however, was challenged on both theoretical and empirical grounds. It has been argued that the processing of A^nB^n grammar does not necessarily require the buildup of a hierarchical structure, but could be based on a simpler computation involving a counting mechanism plus some memory abilities (Perruchet and Rey, 2005; de Vries et al., 2008). The empirical challenge comes from studies reporting that songbirds are able to process A^nB^n grammars (Gentner et al., 2006; Abe and Watanabe, 2011). However, although the grammar used by Abe and Watanabe (2011) can be described as being asymmetric, similar to natural languages, the detection of the incorrect sequences in the experiment could, in principle, be performed based on the following computation: process the incoming sequence of (adjacent) elements and, upon detection of the center element, reverse-and-match the following sequence to the initial sequence (**Figure 2**). The underlying mechanisms used to process symmetrical A^nB^n grammars thus remain speculative, both for the songbird studies (Gentner et al., 2006; Abe and Watanabe, 2011) and for the human study (Friederici et al., 2006).

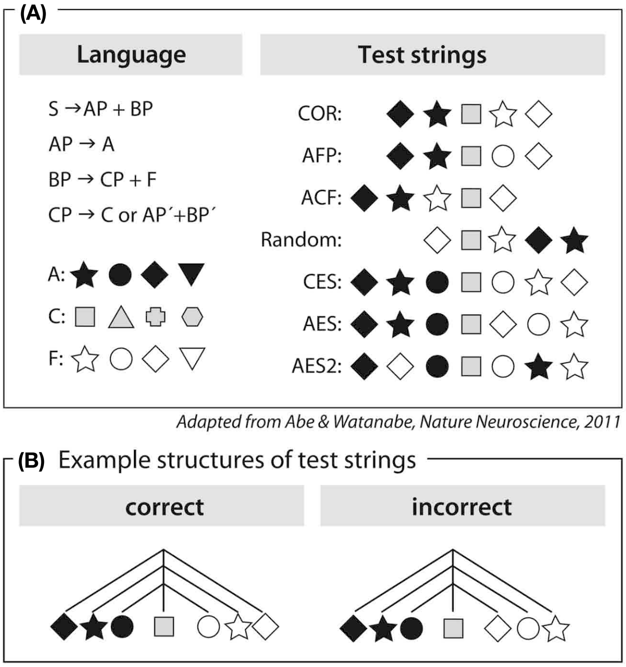
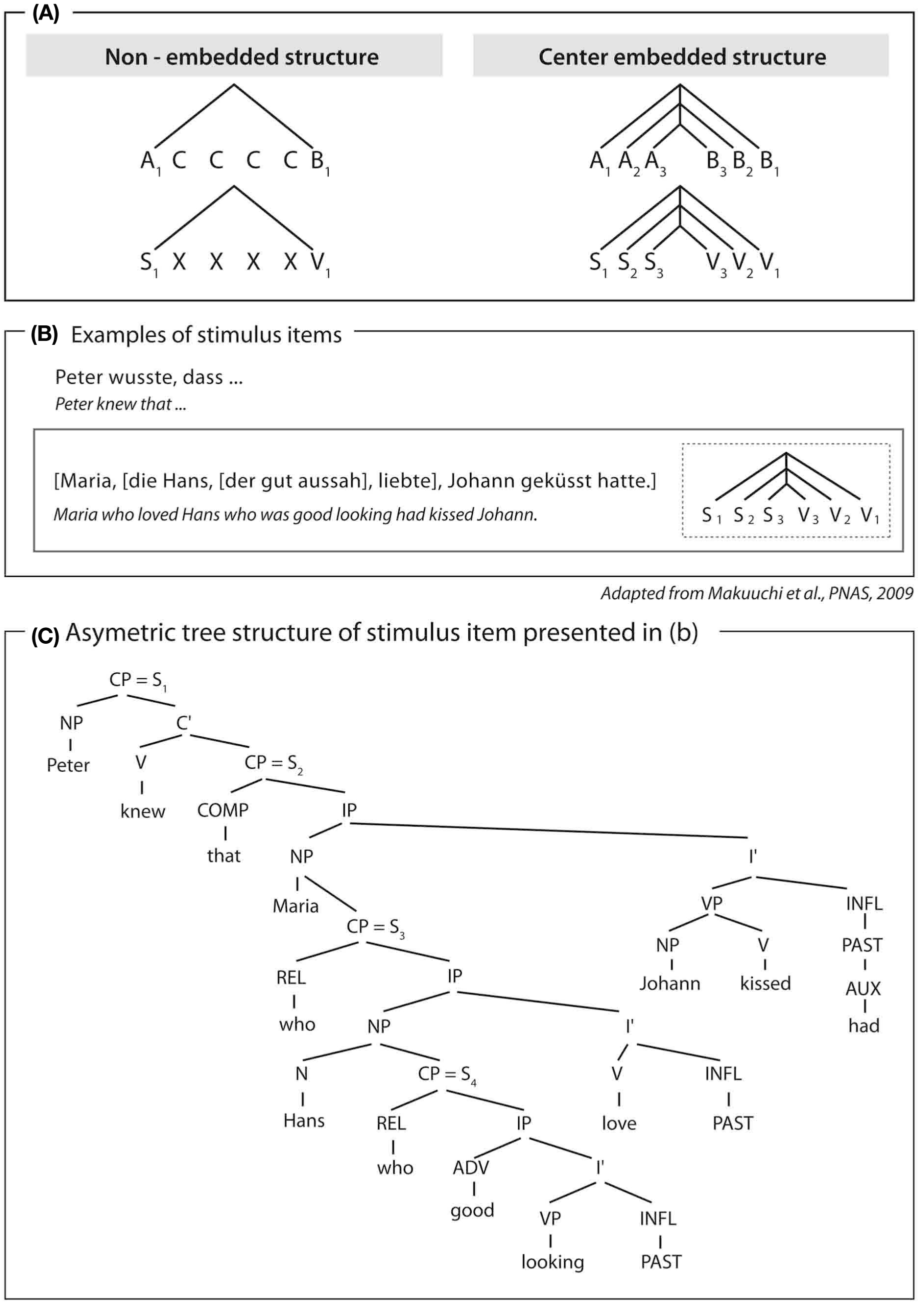


FIGURE 2 | Artificial grammar used in Abe and Watanabe (2011). (A) Left panel: description of the grammar as a center-embedded structure and members of categories A, C, and F; Right panel: examples of test strings. **(B)** Symmetrical description of center-embedded structure (CES) and violating test string (AES). The center element is represented in gray.

For humans, however, the argument can be made that the computation they apply to process symmetrical A^nB^n grammars does indeed involve hierarchy building. The argument is based on two findings. Humans process symmetrical grammatical structures lacking functional categories (Friederici et al., 2006; Bahlmann et al., 2008) and also asymmetrical natural grammars, which require the buildup of a multi-level hierarchy, using the same brain area; namely Broca's area (Makuuchi et al., 2009; see **Figure 3**). This brain region is part of the neural network which is dorsally connected to the temporal cortex via the superior longitudinal fascile and the arcuate fascile (Friederici et al., 2006). The finding that humans process symmetrical structures of artificial A^nB^n grammars within the same brain region used to process hierarchical asymmetrical structures in natural language leads to the conclusion that the underlying mechanism for both is that of building hierarchies.



Adapted from Makuuchi et al., PNAS, 2009

FIGURE 3 | Structure and examples of German sentences used in Makuuchi et al. (2009). (A) Schematic view of non-embedded and embedded structures described symmetrically. **(B)** Example sentence of embedded structure. **(C)** Tree structure (asymmetric) for example sentence displayed in **(B)**. Note that although the structure of the sentence displayed in **Figure 3C** could be described schematically as a symmetrical one (**Figures 3A,B**), this is not an adequate description since natural grammars contain functional categories requiring an asymmetrical description (**Figure 3C**).

For songbirds, the argument concerning the mechanism underlying grammar learning is different. In songbirds, the ability to learn grammatical sequences is based on a brain system mediating auditory input-to-vocal output (Fujimoto et al., 2011, for a review, see Bolhuis et al., 2010). This auditory-to-motor circuit, which probably acts in concert with a memory component, may underlie songbirds' ability to learn symmetric hierarchies (Bloomfield et al., 2011). In humans, the dorsal fiber bundle that connects the sensory auditory cortex to the premotor cortex can be viewed as a candidate neural structure of a functionally parallel auditory-to-motor circuit. This structure appears to play a crucial role in phonology-based language learning in humans during early infancy (Berwick et al., 2011).

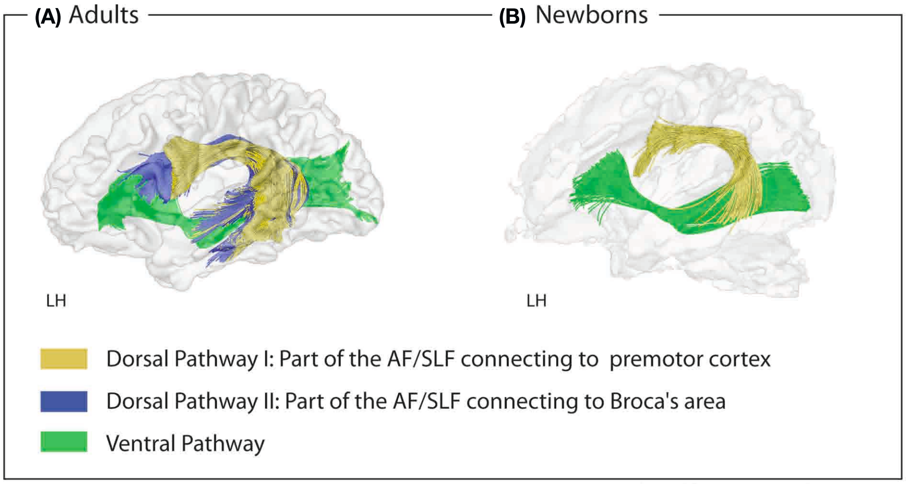
LANGUAGE DEVELOPMENT AND BRAIN MATURATION IN HUMANS

In the past, the dorsal pathway that connects the temporal cortex to the frontal cortex, as observed in adults (Catani et al., 2002), has been proposed to not only support auditory-to-motor mapping (Hickok and Poeppel, 2007; Saur et al., 2008) but also to subserve the processing of syntactically complex sentences (Friederici et al., 2006; Friederici, 2009). Both views are evidenced by data from patient studies. On the one hand, lesions of the dorsal pathway result in conduction aphasia which is characterized by the inability to repeat speech (Geschwind, 1965a,b). On the other hand, lesions of the dorsal pathway correlate with deficits in syntactic processing (Wilson et al., 2011). Unfortunately, however, these patient studies do not allow a functional segregation of different parts of the dorsal pathway. Therefore, ontogenetic data may provide relevant information.

Newborns and infants show impressive language learning abilities. Newborns learn simple grammatical rules from auditory input after brief exposure (Gervain et al., 2008; Teinonen et al., 2009). By the age of 4 months, infants can learn the rule-based dependency of non-adjacent elements in a novel natural language, again after brief exposure to correct sentences (Friederici et al., 2011). During production, very young infants demonstrate a language-specific prosody in their cry patterns, long before they start to babble (Mampe et al., 2009), and during the babbling phase, they continuously tune their production toward the phonology of their target language (de Boysson-Bardies et al., 1984). This early phonology-based learning stage should require a circuit allowing auditory-to-motor mapping. Structural imaging data shows that newborns display such an auditory-to-motor circuit in the form of a dorsal pathway which links the temporal cortex to the premotor cortex (Perani et al., 2011). This dorsal pathway connecting to the premotor cortex must be separated from an additional dorsal pathway that connects the temporal cortex to Broca's area which is present in adults but not myelinized in infants (see **Figure 4**).

Here, it is proposed that there are two functionally distinct parts of the dorsal pathway (see **Figure 4**): one part connecting the temporal cortex to the premotor cortex (hereafter called Dorsal Pathway I) and a second, more medially located part, connecting the temporal cortex to Broca's area (hereafter called Dorsal Pathway II).

Dorsal Pathway I, supporting sensory-to-motor mapping, is present at birth, whereas Dorsal Pathway II is not (Perani et al., 2011). Previous studies with infants between 1 and 4 months old suggested that the dorsal pathway connecting to Broca's area is present early in life (Dubois et al., 2006, 2009), although the data appear to indicate that only the part of the dorsal pathway which connects to the premotor cortex (Dorsal Pathway I) is present. The authors proposed this may be due to methodological problems



Adapted from Perani et al., PNAS, 2011

FIGURE 4 | Fiber tracking of diffusion tensor imaging data with seed in Broca's area and seed in the precentral gyrus/premotor cortex in (A) adults and (B) newborns. Two parts of the dorsal pathway are present in adults; one connecting the temporal cortex via the fasciculus arcuatus (AF) and the superior longitudinal fasciculus (SLF) to the inferior frontal gyrus, i.e., Broca's area (blue), and one connecting the temporal cortex via the AF/SLF to the precentral gyrus, i.e., premotor cortex (yellow). In newborns, only the part connecting to the precentral gyrus can be detected. The ventral pathway connecting the ventral inferior frontal gyrus via the extreme capsule fiber to the temporal cortex (green) is detectable in adults and newborns. LH, left hemisphere.

and base their argumentation on functional data showing that Broca's area is activated in response to speech at this age (Dehaene-Lambertz et al., 2010). The study on newborns, however, indicates that this argumentation must be challenged, since for newborns although they show activation in Broca's area, neither a functional nor a structural connectivity between Broca's area and the temporal cortex can be found (Perani et al., 2011). Thus, it remains to be demonstrated whether the connection between the temporal cortex and Broca's area is fully present in infants between 1 and 4 months old, in particular, since another recent study reports that the dorsal pathway is limited to the parietal and temporal sections in early infancy (Leroy et al., 2011). At least, the data from Perani et al. (2011) indicate that in newborns, only the dorsal pathway that connects to the premotor cortex is myelinized. This connection may not only support tuning processes during babbling, but, moreover, the observed early learning of rule-based dependencies from auditory input in human infants (Gervain et al., 2008; Friederici et al., 2011; Kudo et al., 2011).

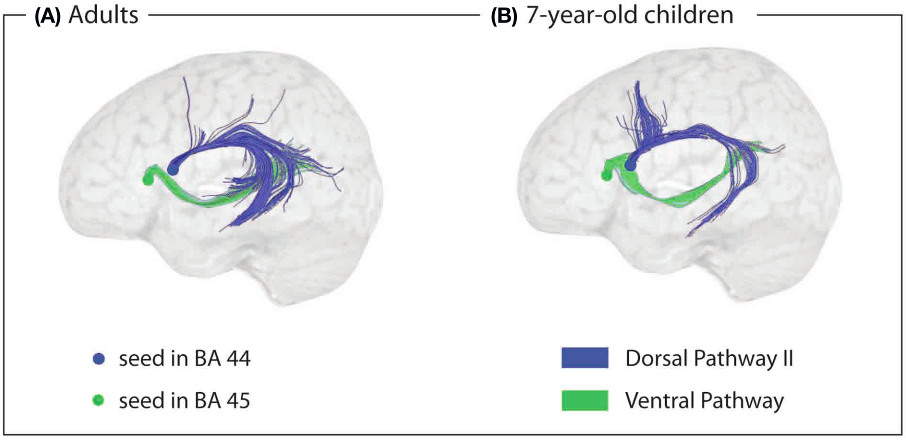
Dorsal Pathway II, connecting the temporal cortex to Broca's area, only develops as the brain matures, and is not even fully myelinized at the age of seven (Brauer et al., 2011; see **Figure 5**). It is argued that Dorsal Pathway II supports the processing of multi-level hierarchically structured sentences. The argument is based on two additional, independent findings. First, adults activate Broca's area and the posterior portion of the superior temporal gyrus and sulcus when processing syntactically complex sentences (Bornkessel et al., 2005; Friederici et al., 2009). Second, only adults with a mature Dorsal Pathway II are able to process syntactically complex sentences correctly, whereas children under the age of seven – an age at which the Dorsal Pathway II is not yet fully matured – are not (Hahne et al., 2004; Dittmar et al., 2008).

These findings in humans make it likely that grammatical rule learning and processing in infants and in adults are partly based on different brain structures. Learning and processing of auditory structured sequences, as shown in infants, could be based on the ability to identify phonological statistical relations of elements in a sequence and some memory capacity. This ability may partly be based on Dorsal Pathway I. In adults, this automatic way of learning from the auditory input is no longer at work, and strategic processes take over (Mueller et al., 2010).

Finally, the question remains: What is the function of the ventral pathway in language processing? The connection between the anterior portion of the prefrontal cortex and the middle temporal gyrus via the extreme capsule fiber system has been related functionally to semantic processing and comprehension (Saur et al., 2008; Tyler and Marslen-Wilson, 2008). There is accumulating evidence in support of the view that this pathway is relevant for semantic processes (for a review, see Weiller et al., 2011). However, the ventral pathway also appears to be involved in the processing of some aspects of syntax (Friederici et al., 2006; Tyler et al., 2011). The data on this issue are sparse, but it seems that the processing of simple sentences (e.g., *The pilot is flying the plane*) does not necessarily recruit the posterior portion of Broca's area, and thus the dorsal pathway (Saur et al., 2008; Tyler and Marslen-Wilson, 2008). The finding that the ventral pathway supports the computation of semantic relations, as well as some syntactic dependencies, raises the question of whether there is a general underlying mechanism capturing both aspects, or whether one has to assume two ventral pathways. Future studies will have to resolve this issue.

CONCLUSION

In light of these across species and within-human findings, we can speculate that there is a parallel mechanism for sequence learning across species, which is based on an auditory (input)-to-motor (output) circuit. In songbirds, the causal relation between the auditory input-to-vocal output and sequence learning is well established (Scharff and Nottebohm, 1991; Fujimoto et al., 2011). In humans, a respective neural circuit – Dorsal Pathway I connecting the temporal cortex to the premotor cortex – is present at birth and may be responsible for rule-based sequence learning observed in young infants (Teinonen et al., 2009; Kudo et al., 2011). In adults,



Adapted from Brauer et al., Cerebral Cortex, 2011

FIGURE 5 | Fiber tracking of diffusion tensor imaging data with seeds in Brodmann Area (BA) 44 and 45 in (A) adults and (B) 7-year-old children. The dorsal pathway connects the posterior part of Broca's area (BA 44) to the temporal cortex via the AF/SLF. The ventral pathway connects the anterior part of Broca's area (BA 45) to the temporal cortex via the extreme capsule fiber system.

this neural network appears to support bottom-up processes such as speech perception and repetition (Saur et al., 2008).

However, Dorsal Pathway II, connecting Broca's area to the temporal cortex, may specifically subserve the processing of language-like hierarchical structures. The supporting evidence for this is twofold: first, the dorsal pathway is generally stronger in human adults than in non-human primates (Rilling et al., 2008), and second, Dorsal Pathway II, in particular, is not myelinized at birth and only fully develops at around the time children master syntactically complex sentences (Brauer et al., 2011; Perani et al., 2011). In adults, this pathway may be involved in fronto-to-temporal top-down predictive processes during language comprehension.

ACKNOWLEDGMENTS

I thank the two reviewers and Jens Brauer for helpful comments on the manuscript.

REFERENCES

Abe, K., and Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* 14, 1067–1074.

Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., and Knösche, T. R. (2007). Connectivity-based parcellation of Broca's area. *Cereb. Cortex* 17, 816–825.

Bahlmann, J., Schubotz, R., and Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *NeuroImage* 42, 525–534.

Berwick, R. C., Okanoya, K., Beckers, G. J. L., and Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci. (Regul. Ed.)* 15, 113–121.

Bloomfield, T. C., Gentner, T. Q., and Margoliash, T. (2011). What birds have to say about language. *Nat. Neurosci.* 14, 947–948.

Bolhuis, J. J., Okanoya, K., and Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* 11, 747–759.

Bornkessel, I., Zyssett, S., Friederici, A. D., von Cramon, D. Y., and Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage* 26, 221–233.

Brauer, J., Anwander, A., and Friederici, A. D. (2011). Neuroanatomical prerequisites for language functions in the maturing brain. *Cereb. Cortex* 21, 459–466.

Catani, M., Howard, R. J., Pajevic, S., and Jones, D. K. (2002). Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage* 17, 77–94.

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

de Boysson-Bardies, B., Sagart, L., and Durant, C. (1984). Discernible differences in the babbling of infants according to target language. *J. Child Lang.* 11, 1–15.

de Vries, M. H., Monaghan, P., Knecht, S., and Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: the learnability of embedded hierarchical structures. *Cognition* 107, 763–774.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., and Dehaene, S. (2010). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14240–14245.

Dittmar, M., Abbot-Smith, K., Lieven, E., and Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Dev.* 79, 1152–1167.

Dubois, J., Hertz-Pannier, L., Cachia, A., Mangin, J. F., Le Bihan, D., and Dehaene-Lambertz, G. (2009). Structural asymmetries in the infant language and sensorimotor networks. *Cereb. Cortex* 19, 414–423.

Dubois, J., Hertz-Pannier, L., Dehaene-Lambertz, G., Cointepas, Y., and Le Bihana, D. (2006). Assessment of the early organization and maturation of infants' cerebral white matter fiber bundles: a feasibility study using quantitative diffusion tensor imaging and tractography. *Neuroimage* 30, 1121–1132.

Fitch, W. T., and Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends Cogn. Sci. (Regul. Ed.)* 13, 175–181.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., and Anwander, A. (2006). The brain differentiates human, and non-human grammars: functional localization, and structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2458–2463.

Friederici, A. D., Makuuchi, M., and Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *Neuroreport* 20, 563–568.

Friederici, A. D., Müller, J., and Oberecker, R. (2011). Precursors to natural grammar learning: preliminary evidence from 4-month-old infants. *PLoS ONE* 6, e17920. doi:10.1371/journal.pone.0017920

Fujimoto, H., Hasegawa, T., and Watanabe, D. (2011). Neural coding of syntactic structure in learned vocalizations in the songbird. *J. Neurosci.* 31, 10023–10033.

Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.

Gervain, J., Macagno, F., Cogoi, S., Pen, A. M., and Mehler, J. (2008). The neonate brain detects speech structure. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14222–14227.

Geschwind, N. (1965a). Disconnexion syndromes in animals and man. I. *Brain* 88, 237–294.

Geschwind, N. (1965b). Disconnexion syndromes in animals and man. II. *Brain* 88, 585–644.

Hahne, A., Eckstein, K., and Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *J. Cogn. Neurosci.* 16, 1302–1318.

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

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech perception. *Nat. Rev. Neurosci.* 8, 393–402.

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., and Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Dev. Sci.* 14, 1100–1106.

Leroy, F., Glasel, H., Dubois, J., Hertz-Pannier, L., Thirion, B., Mangin, J.-F., and Dehaene-Lambertz, G. (2011). Early maturation of the linguistic dorsal pathway in human infants. *J. Neurosci.* 31, 1500–1506.

Makris, N., and Pandya, D. (2009). The extreme capsule in humans and rethinking of the language circuitry. *Brain Struct. Funct.* 213, 343–358.

Makuuchi, M., Bahlmann, J., Anwander, A., and Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8362–8367.

Mampe, B., Friederici, A. D., Christophe, A., and Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Curr. Biol.* 19, 1994–1997.

Mueller, J. L., Bahlmann, J., and Friederici, A. D. (2010). Learnability of embedded syntactic structures depends on prosodic cues. *Cogn. Sci.* 34, 338–349.

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., and Friederici, A. D. (2011). Neural language networks at birth. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16056–16061.

Perruchet, P., and Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychon. Bull. Rev.* 12, 307–313.

Petrides, M., and Pandya, D. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7, e1000170. doi:10.1371/journal.pbio.1000170

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X. Y., Zhao, T. J., Hu, X. P., and Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M. S., Umarova, R. M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., and Weiller, C. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040.

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.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., and Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 10, 21. doi:10.1186/1471-2202-10-21

Tyler, L. K., and Marslen-Wilson, W. D. (2008). Fronto-temporal brain systems supporting spoken language comprehension. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1037–1054.

Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., Papoutsi, and M., and Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: function, structure and behaviour in left-hemisphere damaged patients. *Brain* 134, 415–431.

Weiller, C., Bormann, T., Saur, D., Musso, M., and Rijntjes, M. (2011). How the ventral pathway got lost: and what its recovery might mean. *Brain Lang.* 118, 29–39.

Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., Ogar, J. M., DeLeon, J., Miller, B. L., and Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron* 72, 397–403.

Conflict of Interest Statement: The author declares 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: 06 October 2011; accepted: 18 January 2012; published online: 06 February 2012.

Citation: Friederici AD (2012) Language development and the ontogeny of the dorsal pathway. Front. Evol. Neurosci. ***4****:3. doi: 10.3389/fnevo.2012.00003*

frontiers in
EVOLUTIONARY NEUROSCIENCE

REVIEW ARTICLE
published: 03 July 2012
doi: 10.3389/fnevo.2012.00009

The dual loop model: its relation to language and other modalities

Michel Rijntjes, Cornelius Weiller*, Tobias Bormann and Mariacristina Musso

Neurologische Universitätsklinik Freiburg, Freiburg, Germany

Edited by:
Angela D. Friederici, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Michel Thiebaut De Schotten, King's College London, UK
Alfred Anwander, Max Planck Society, Germany

***Correspondence:**
Cornelius Weiller, Neurologische Universitätsklinik Freiburg, Breisacher Strasse 64, D-79106 Freiburg, Germany.
e-mail: cornelius.weiller@uniklinik-freiburg.de

The current neurobiological consensus of a general dual loop system scaffolding human and primate brains gives evidence that the dorsal and ventral connections subserve similar functions, independent of the modality and species. However, most current commentators agree that although bees dance and chimpanzees grunt, these systems of communication differ qualitatively from human language. So why is language unique to humans? We discuss anatomical differences between humans and other animals, the meaning of lesion studies in patients, the role of inner speech, and compare functional imaging studies in language with other modalities in respect to the dual loop model. These aspects might be helpful for understanding what kind of biological system the language faculty is, and how it relates to other systems in our own species and others.

Keywords: language, DTI, dual loop model

INTRODUCTION

Current cognitive, neuropsychological and neurobiological theories assume that a dual system scaffolds the organization of the brain. Primate models initially showed that two parallel pathways, an anterolateral or ventral and a caudolateral or dorsal, interact primary with non-primary visual (Mishkin et al., 1983; Kastner and Ungerleider, 2000) or acoustic cortex (Romanski et al., 1999a; Rauschecker and Tian, 2000; Tian et al., 2001). Neuropsychological (Clarke et al., 2000, 2002) and neuroimaging (Ahveninen et al., 2006; Bernal and Ardila, 2009) studies have at first discussed both these pathways also in human visual or acoustic systems and then the model has been extended to the motor system (Rizzolatti and Matelli, 2003) and to higher cognitive functions such as attention (Corbetta et al., 2005) and language (Paulesu et al., 2003; Demonet et al., 2005; Hickok and Poeppel, 2007). A recent review on connectivity of the prefrontal cortex in monkeys confirms the dorsal-ventral dichotomy of projections of long association connections to post-rolandic regions (Yeterian et al., 2012).

It is a novel method, diffusion-tensor-imaging (DTI) based fiber tracking, that finally makes it possible *in vivo* to identify long human association tracts for ventral and dorsal pathways, similar to animal data (Kreher et al., 2008). The extreme capsule (EmC) and uncinate fascicle (UF) are part of the ventral system, and the superior longitudinal fasciculi (SLF) (SLFI, II, III) and the arcuate fasciculus (AF) are all dorsal pathways (Makris et al., 1999; Wise, 2003; Parker et al., 2005; Anwander et al., 2007; Frey et al., 2008; Thiebaut de Schotten et al., 2011) (See **Box 1**). Using DTI, it is possible to correlate probabilistic tracking to functional imaging results, relating functionally defined ventral or dorsal pathways to specific tasks and modalities, thus identifying the possible functional role of the underlying pathways (Saur et al., 2010). Obviously, this has particular advantage for those higher faculties that are absent in animals.

Actual data seem to confirm only to some extent initial speculations based on anatomical or functional imaging data. For example, in the attention system, Umarova et al. showed that in addition to dorsal connections, an interaction between temporo-parietal cortex, anterior insula, and inferior frontal gyrus takes place along the ventral network (Umarova et al., 2010), refuting previous assumptions (Corbetta et al., 2005). In contrast to our textbook perception, a dual pathway model (as initially already discussed by Wernicke) is currently also accepted for language (Weiller et al., 2011), with few exceptions (Catani et al., 2005; Ross, 2010). Thus, current neurobiological consensus of a broad dual system scaffolding human and lower primate brains gives evidence that the dorsal and ventral connections subserve similar functions, independent of the modality (Weiller et al., 2011), even if the post- and pre-rolandic areas involved might be different and still modality-dependent.

However, a main difference between humans and lower primates is our unparalleled sophistication in communication capacity. We are not only able to describe objects concretely in words that can be understood by others, but we are also capable of connecting acoustic utterances (words) with very abstract connotations and emotions ("to be or not to be," "yes we can"). But if language is based in principle on the same dual loop system used for communication in lower primates and this system is also found in other modalities (as perception or attention), why is language unique to humans? We will look at four aspects that are related to the dual loop system, and then ask the question again how unique language is.

Box 1

The EmC is a tract localized between the claustrum and the insular cortex present in monkeys, macaque as well in humans. Fibers running through the EmC connect the superior temporal sulcus (area TPO), the superior temporal gyrus (T1a, paAlt, TAa, TS3), the planum temporale, the rostral insula and in humans also the middle temporal region (TS2), inferior temporal region (i.e., area TE) with the frontal lobe—primarily with area 45, in monkey with area 45A (Petrides and Pandya, 2009), in human pars triangularis (Makris and Pandya, 2009), frontal operculum (FO), pars orbitalis (area 47) and with a modest contingent of fibers concluding in the pars opercularis (area 44) and in the dorsolateral cortex (9/46) and ventral area 10 (Petrides and Pandya, 1988, 2007, 2009; Schmahmann and Pandya, 2006). The IPL—PF, PG, and PFG in monkeys, corresponding to the SMG and angular gyrus in humans, respectively—are connected to the ventrolateral prefrontal and the superior and middle temporal cortex via the middle longitudinal fasciculus MDLF (Seltzer and Pandya, 1984), while some parts of the IPL are probably connected via the EmC to the inferior frontal lobe (Berke, 1960; Caspers et al., 2011).

Makris et al. showed that the EmC in human directly reaches the IPL (Makris and Pandya, 2009). Thus, EmC, which in part overlaps with tracts called fasciculus occipito-frontalis inferior (Gloor, 1997), inferior occipital fascicle (Duffau et al., 2009) or IFOF (Catani et al., 2002; Thiebaut de Schotten et al., 2012), allows the integration of information between the temporal, the parietal and insular cortex with prefrontal cortices (Bucy and Kluver, 1955; Gloor, 1997; Yeterian et al., 2012).

The dorsal system is generally subdivided in SLF connecting the angular gyrus (or the area PG in monkeys), rostral part of the SMG (equivalent to primate area PF) and caudal part of the suparamarginal gyrus (corresponding to are PFG) with Broca's region in the frontal lobe (BA 44 and 45) and the surrounding dorsal and ventral areas 6 and 9/46 (Schmahmann and Pandya, 2006; Petrides and Pandya, 2009). The arcuate fasciculus connects the adjacent superior temporal sulcus (Catani et al., 2005) or middle temporal gyrus and prefrontal regions.

Numerous methodological problems in DTI-based fiber tracking remain, like lack of quantification, limited spatial resolution and the problem of crossing fibers. Also, it is unclear whether particularly the ventral route along EmC is monosynaptic or polysynaptic with obligatory interruption in the insula and claustrum (Makris and Pandya, 2009; Petrides and Pandya, 2009; Weiller et al., 2011). The number of pathways within the dorsal and ventral systems, their exact origin, their course and their endings, and therefore their exact functions, are still debated. All these questions demand further development of tracking techniques.

WHAT IS SPECIAL IN THE HUMAN DUAL LOOP MODEL?

To what extent does language share the same anatomical systems as lower primates and what are the differences? Several explanations are possible, and they need not be mutually exclusive. The first possibility (A) is that there is an evolutionary change in a specific pathway (ventral or dorsal). Another possibility (B) is the hypothesis of an anatomical-functional gradient along the dual system, meaning the additional development of pre- and post-rolandic modules located more anteriorly and posteriorly, functionally enabling a new level of interaction between dorsal and ventral pathways.

Anatomical innovations are generally discussed in an evolutionary framework of a serial evolution of humans from non-human primates. However, we will discuss that basic rules of brain organization, which humans still share not only with non-human primates, but also with cetaceans and songbirds, could also lead to a parallel and separate evolution (C).

THE NEUROANATOMICAL DIFFERENCES OF A SPECIFIC PATHWAY BETWEEN HUMAN AND NON-HUMAN SPECIES MAY BE CRUCIAL FOR THE EVOLUTION OF LANGUAGE

The location of the EmC within the language zone, specifically connecting Broca's area in the frontal lobe and Wernicke's area in the temporal lobe and inferior parietal lobule, suggests that the EmC may have the prominent role for what Mesulam has called the language communication epicenter (Mesulam, 1998; Schmahmann and Pandya, 2006; Makris and Pandya, 2009; Weiller et al., 2009) and Meynert the "central language complex" (Meynert, 1866; Weiller et al., 2011). Indeed, a recent study of our lab showed that "mapping sound to meaning" crucially relates to the interaction between Broca's and Wernicke's area along connections running through the EmC (Saur et al., 2008, 2010) and electrical stimulation of the anterior floor of the EmC, corresponding to parts of the inferior fronto-occipital fascicle (IFOF), can generate semantic paraphasias (Duffau et al., 2005; Duffau, 2012).

Another, more actively discussed theory proposes that the dorsal pathway projecting from the posterior portion of Broca's area to the superior temporal region—the AF—seems to be of particular importance for language as it is involved in word repetition (Saur et al., 2008) and especially in phrase-structure grammar (Friederici, 2009, 2011). Non-human primates (as well newborns) may not be able to process hierarchically acoustic sequences as well as to repeat complex acoustic signals (Hauser et al., 2001). Fiber tracking studies in chimpanzees, macaques (*Macaca mulatta*) and humans reveal that particularly the AF within the dorsal pathways, even if undoubtedly present (Petrides and Pandya, 2009), became more prominent during the evolution (Rilling et al., 2008). In a recent comparative study of human and monkey association tracts of the frontal lobe, many similarities were present, but one major difference was found in the AF, with the majority of fibers from this tract projecting to the middle and inferior temporal gyri in human but not in monkey (Thiebaut de Schotten et al., 2012). Moreover, the white matter tracts in the cella media show an asymmetry favoring the left side in the degree of anatomical connectivity by microscopic examination of post-mortem specimens (Galuske et al., 2000), by structural T1 MRI (Paus et al., 1999) and by DTI (Buchel et al., 2004; Nucifora et al., 2005; Hagmann et al., 2006; Powell et al., 2006; Catani et al., 2007), even if further studies are clearly needed to establish the functional-anatomic relationship with respect to lateralization (Hagmann et al., 2006). Structural studies indicate that the dorsal pathway is weaker in children compared to adults (Zhang et al., 2007; Dubois et al., 2008; Lebel et al., 2008) and matures only after the age of seven (Brauer et al., 2011). Given the function of the dorsal system, i.e., the integration of forward and inverse models—from sensorimotor integration involved in language repetition (Saur et al., 2008), to overt or covert speech

(Oppenheim and Dell, 2008) to integrative processing of long-time dependence (Friederici et al., 2006), it could be concluded that all these current tracking data indicate that the sophistication of this competence may be essential for the evolution of the language faculty and for human uniqueness in general.

However, whether the thickness of a pathway is the most determinate argument of human evolution is actually unclear. Thickness may be influenced by genetic selection and may depend on practice. In a genetically defined disease, decreased regional anisotropy of the left AF was found in children with Angelman syndrome, pointing to a possible relation between the AF, some aspect of language (i.e., production) and genetic (innate) constraints (Peters et al., 2011; Wilson et al., 2011). However, decreased fractional anisotropy in Angelman syndrome did not selectively involve the AF, but also the inferior fronto-occipital fasciculus, cingulum, anterior thalamic and brainstem radiation, as well as the uncinate fasciculus, suggesting that the loss of UBE3A gene expression may result in a widespread abnormal brain connectivity (Tiwari et al., 2012). Also, more is not always better: increased fractional anisotropy of the superior longitudinal fasciculus is associated with poor visuospatial abilities in Williams Syndrome (Hoeft et al., 2007), another neurodevelopmental disorder, and volumetric increase of arcuate projections are observed in autism (Casanova et al., 2010). Moreover, also practice can influence the size of a pathway. Professional musicians show an increased size of the right AF in relation to the degree of musical expertise (Oechslin et al., 2009), and melodic intonation therapy may induce an increase of fractional anisotropy values along the AF (Schlaug et al., 2009), suggesting a use dependency of the size of the pathway. Continuous speaking may determine the size of the AF, and humans are not only able to chatter but they do it incessantly. Thus, the higher volume of arcuate projections in adults in comparison to children and lower primates may not represent the main argument for human uniqueness of the language faculty.

THE HYPOTHESIS OF AN ANATOMICAL-FUNCTIONAL GRADIENT ALONG THE DUAL SYSTEM

The ability to process syntactically complex rules for sure represents a core component of the human language faculty, as it allows its richness of expressivity that is lacking in the animal communication system, "where each sound is associated with a particular meaning but sounds are not recombined to form a new meaning" (Patel, 2008). However, till now, more detailed studies using combinations of fMRI and DTI are needed to explore this aspect in natural language. In Saur's experiment simple sentences were compared with reversed speech and, therefore, did not give any information about the ability of recursion in language (Saur et al., 2008). Even the capacity to classify items that have already been instantiated in a given pattern into simple phrase-structure sequences as in Friederici's artificial grammar task (Friederici et al., 2006) has to be differentiated from recursion. Indeed, the ability to recognize acoustic patterns defined by a self-embedding, context-free grammar, even if it seems not to be present in non-human primates, at least for the acoustic modality (Hauser et al., 2001), is not unique to humans, but also exists in songbirds (Gentner et al., 2006). Moreover, in Friederici's experiment, processing a more complex grammar involved not only dorsal but also ventral pathways, while more simple finite-state grammar relates only to the ventral one. Grammar complexity may be a factor to differentiate the ventral from the dorsal system (Friederici et al., 2006). Alternatively, processing of local as well as non-adjacent dependencies point to a time-independent analysis, and this aspect is processed along the ventral network (Belin and Zatorre, 2000; Rauschecker and Scott, 2009). Processing of long-distant dependencies could additionally necessitate a time-dependent analysis, which requires a continuous integration of feed-forward and inverse models and thus may be principally relate to the dorsal pathway (Belin and Zatorre, 2000; Rauschecker and Scott, 2009; Weiller et al., 2011). Such functional differentiation—i.e., between time-dependent and time-independent analysis (for further discussion, See section "Comparing The Dual Loop Model In Language With Other Modalities: What Are The Essential Characteristics Of The Ventral And Dorsal Pathways?")—is also present in monkeys (Zuberbuhler, 2002). The pivotal element for evolution, therefore, as discussed in section "The Neuroanatomical Differences Of A Specific Pathway Between Human And Non-Human Species May Be Crucial For The Evolution Of Language," may be not the volumetric increase of the projections of a specific pathway, but the higher cellular differentiation of the "terminal" brain regions connected by the "language dual system." This enables enhanced connectivity of dorsal and ventral pathways and thus simultaneous processing between the two pathways, for which increasing evidence exists, not only in the language modality (Rosazza et al., 2009; Rolheiser et al., 2011), but also in the acoustic (Leavitt et al., 2011), visuospatial (Almeida et al., 2010) and visuomotor modality (Creem and Proffitt, 2001; Mahon et al., 2007).

In the course of the last 50 years or so, in addition to Wernicke's and Broca's area, other regions as supramarginal gyrus (SMG), inferior parietal lobe (IPL), angular gyrus, anterior temporal lobe have been identified to be fundamental for language processing. Although the brains of chimpanzee and macaque,—but also of bonobo, gorilla, and orangutan—show a human-like left-right-hemisphere asymmetry, cellular and functional heterogeneity and similar anatomical connectivity through dorsal and ventral systems (Cantalupo and Hopkins, 2001; Schenker et al., 2008), they possess a smaller (in term of "gyral" white matter) frontopolar cortex and smaller middle temporo-parietal regions (Schenker et al., 2005). At the microscopic level, in Broca's area (Schenker et al., 2008) and in the planum temporale (Anderson et al., 1999; Buxhoeveden et al., 2001a,b), a stronger left-right asymmetry (in terms of cellular density) and a greater horizontal spacing distance are observed in humans. The increased horizontal spacing in humans reflects an increased number of input and output connections and an increased microcircuitry. Indeed, comparative studies on structural connectivity found that in humans, tract terminations in middle and inferior temporal Gyrus, as well as in pars opercularis (Brodmann Area (BA) 44), pars triangularis (BA 45), pars orbitalis (BA 47) of the inferior frontal gyrus and surrounding regions are much stronger than in macaques or chimpanzees (Rilling et al., 2008; Petrides and Pandya, 2009). Ventral interactions between inferior parietal and inferior frontal lobe are reported in old anatomical (Berke, 1960) and DTI-based fiber tracing studies in humans (Saur et al., 2008; Makris and Pandya, 2009; Umarova et al., 2010; Vry et al., 2012). It is still an unsettled question, however, whether these reflect monosynaptic or polysynaptic pathways (Weiller et al., 2011).

All these data may suggest that in a system of two equivalent pathways, hierarchy is not determined by one specific pathway, although specific functions may primarily or crucially involve one of both, but rather by an extension of this system to regulatory, cytoarchitectonically more developed areas in prefrontal, temporal and parietal neocortex in humans (Weiller et al., 2011). The caudal extension along the temporal lobe is observed by processing tones and noise bursts into words and sentences (Rauschecker and Scott, 2009), while the additional involvement of even more caudal temporo-parietal areas (comprising Wernicke's area, angular and inferior parietal gyrus) is associated with increasing semantic complexity (Sharp et al., 2010). A gradient along the frontal and prefrontal cortex is thus linked to increasing abstraction of the underlying processes (Badre and D'Esposito, 2009). In a series of DTI studies from our lab, it seems that the terminations of the ventral system may be more anteriorly and posteriorly than the respective terminations of the dorsal system, in the sense that the ventral system seems to "embrace" the dorsal system (**Figure 1**). Additionally, an increase in connectivity within and between the cortical regions representing the terminal parts of the dual system could also induce a closer interaction between both dorsal and ventral systems and within each system. It is suggested that such an interaction could especially take place in the prefrontal cortex, with local connections permitting functional interactions of processes in dorsal and ventral pathways (Yeterian et al., 2012).

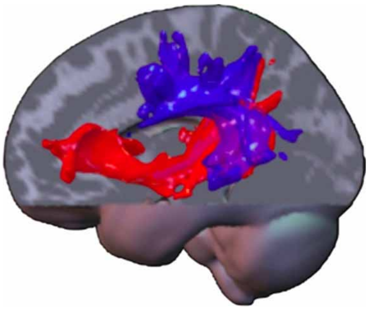


FIGURE 1 | A composite display of tracking-related to dorsal and ventral connections from the different studies of our lab in various domains and modalities illustrates commonalities and differences. The ventral tracts have a wider radius, "embracing" the dorsal ones. Note that trackings between parietal and temporal lobe contain dorsal and ventral pathway-related fibers. The more laterally located dorsal pathway-related fibers may constitute the ascending limb of the arcuate fasciculus. The ventral pathway-related fibers may either use the MdlF for parieto-temporal exchange or a potential "parietal part" of the extreme capsule, which connects parietal cortex with prefrontal cortex (Makris and Pandya, 2009). This tract already displayed on the frozen sections of Ludwig and Klingler (Ludwig and Klinger, 1956). Fiber tracts within the temporal lobe may be related to the MdlF, before aligning with those from the parietal cortex and entering the extreme capsule for the prefrontal cortex, potentially being identical with the anterior part of what is called the IFOF. Note, anatomy is a vehicle for pathways but not identical and assumed fiber location derived from probabilistic tracking may be part of defined strong white matter tracts or not (like AF; SLF, IFOF). The latter ones run through anatomically defined regions, which contain mainly white matter and may be constituted of different long (and short) tracts as the cella media (containing the AF, SLF system) or the extreme capsule. Note: most association connections are reciprocal.

According to this view, the fronto-temporal interactions envelop the insula, the claustrum and the basal ganglia, anatomical structures with projections to almost all cortical regions (Edelstein and Denaro, 2004; Schmahmann and Pandya, 2006; Mathur et al., 2009). The exact role of these regions in language continues to be discussed, but at a very general level there is an agreement that they may carry out mainly integrative functions, enabling the information from one modality with information from other modalities: The basal ganglia seem to play a crucial "integration" role particularly along the dorsal system (Rauschecker and Scott, 2009). In the ventral system, the insula might be responsible for the integration of auditory information with other associative functions (Bamiou et al., 2003), or integration of basic information (features) for initial categorization or grouping (Bamiou et al., 2006). The claustrum seems to bind sensory inputs within and across sensory modalities to generate conscious percepts (Crick and Koch, 2005). Interestingly, Meynert already considered the region including the claustrum, insula and the ascending acoustic fibers in the external and EmC as the central language complex (Meynert, 1866).

HUMAN INDEPENDENT EVOLUTION THROUGH BASIC UNIVERSAL RULES OF CORTICAL DEVELOPMENT FOR BRAIN FUNCTIONAL AND ANATOMICAL SEGREGATION

Despite the deep evolutionary divergence between animals, adaptation to physically dissimilar environments, and very different neuroanatomical organization, accumulating evidence indicates that in all regions of the neocortex in humans and in all other mammalian species thus far evaluated, including dolphins (Mountcastle, 1957, 1978; Purves et al., 1992; Krubitzer, 1995; Manger et al., 1998; Buxhoeveden and Casanova, 2002; Butti and Hof, 2010; Casanova et al., 2010), but also songbirds [African gray parrots (Pepperberg and Shive, 2001) and starlings (Gentner et al., 2006)], the smallest level of vertical and horizontal organization in the cortex consists of cells assembled in minicolumns. Minicolumns represent the basic architectonic and physiological elements by which the neocortex organizes its myriad number of neurons (with various specializations), its pathways and intrinsic circuits into a coherent functional unit (Mountcastle, 1957, 1978; Szentagothai, 1983; Manger et al., 1998; Casanova et al., 2006). Evolution across species is generally related to an increase in the number of radial columnar units without significantly changing the number of neurons within each unit (Rakic, 1995). Indeed, despite the difference in cortex expansion, the size of modules in primates or cetaceans is similar to that described for small-brained mammals like the mouse, suggesting that module size is evolutionarily stable across species (Horton and Hedley-Whyte, 1984; Livingstone and Hubel, 1984; Tootell et al., 1996; Manger et al., 1998). "The ubiquity of modules and the apparent convergent evolution of module size" in primates, cetaceans, and carnivores, animals that are separated by as much as 130 million years of independent evolution, indicate that there may be underlying homologous rules of cortical development that cause "initial segregations" (Manger et al., 1998). One of the possible rules could be the "component placement optimization": The length of cortico-cortical connections plays a key role in determining module size (Cherniak, 1994): the shorter the connections the more efficient the processing (Ringo, 1991; Szymanski et al., 1995). In humans, the number of modules increases with as a consequence longer association tracts connecting more distant modules in different areas. For the evolution of biological life, rapid reaction time to sensory input is crucial. Therefore, modules of a limited size are conserved also in an animal whose neocortex has undergone a huge expansion, where long cortico-cortical connections would require a large volume of metabolically active tissue and an increased conduction time (Horton and Hedley-Whyte, 1984; Tootell et al., 1996). The restricted range in size of modules having diverse cytoarchitectonic or histochemical features across highly divergent mammalian species may reflect "an independent evolution possibly due to selection for an optimal connection length" (Manger et al., 1998).

"Component placement optimization" and its balance with evolutionary constant increase of the minicolumns could explain why short connections, dorsal along the SLF as well ventral along EmC, are constant in primate evolution. The optimization of connective processes within minicolumns could be one of the fundamental criterions to adapt minicolumns in the various cortical areas according to their specific developmental and functional requirements. Species-specific differences of modular organization mainly regard the horizontal minicolumn spacing and the resulting numbers of the input and output pathways. In chimpanzees, high density modules localized in primary visual cortex (area 17) seem to be crucial to process color, form and motion; in dolphins, cell clusters in insula (Area 1) may subserve the processing of complex auditory stimuli associated with dolphin communication. So, a more extensive parcellation and the resulting increased hierarchical organization of modular cortical subdivisions of species-specific brain regions seems to led to functional optimization. Even if parcellation can theoretically be found throughout the cortex regardless of functional attributes, it selectively affects brain regions that are known to be functionally specialized.

Thus, it is maybe no coincidence that in humans, the presence of larger minicolumns spacing, as far as is known until now, is characteristic only of language-related cortical areas: in Broca's area (Amunts et al., 1999) and in the planum temporale (Anderson et al., 1999). No asymmetries were found in these two regions in chimpanzees and other primates (Sherwood et al., 2007). A recent tracking study showed not only a cytoarchitectonic, but also a tractographic parcellation within Broca's area (Anwander et al., 2007). Williams et al. made the step from micro-to macro-organization of the brain. They showed that compared to control subjects, autistic persons exhibit a reduced minicolumnar width and peripheral neuropil spacing, as well as an increased number of minicolumns (Williams and Casanova, 2010). This correlates with an increment of short connections (coming from each single minicolumn), which could be related to the highly efficient analytic processing sometimes observed in savants, and a decrease of longer connectivity (particularly of AF and cingulum bundle), which may relate to a deficit in the identification of relationships.

Both humans and non-human primates have commonalities in anatomy, and the capability of language in humans is often put in relation to these differences in anatomy, e.g., a larger AF in humans, as a "next step" in evolution, enabling e.g., complex grammar. However, humans did not evolve from the apes but we evolved with the apes from a common ancestor. As we discussed above, similarities of modular arrangements between homo sapiens and animals are not necessarily the product of the same development, but basic brain rules of brain organization can lead to new functions, even if emerging from similar structures (Northcutt and Kaas, 1995; Vates and Nottebohm, 1995). This could limit the significance of comparative studies on anatomical data in humans and non-human primates.

CAN HUMAN ANATOMY OF THE DUAL LOOP MODEL BE RECONCILED WITH COGNITIVE MODELS OF SPEECH?

The different perspectives of cognitive models and (anatomy-based) neurosciences seem to converge on a "two route model." The famous "house model" of the early aphasiologists derived from patients studies [e.g., Lichtheim and Wernicke (Weiller et al., 2011)] has remained the basic framework for later models of single word processing, for example, Morton's Logogen model (Morton and Patterson, 1980). The latter remains the reference for cognitive studies of aphasic individuals. Cognitive models of speech processing and word production have, on the other hand, included evidence from psycholinguistic studies. All models contain representations both of phonological segments and of conceptual knowledge (**Table 1**). They differ, however, with regard to the number of levels, which mediate between phonological and conceptual information, i.e., whether processing is fully interactive or whether interaction is restricted. In the model of Levelt et al. two lexical levels mediate between semantics and phonology while the interaction between these levels is limited (Levelt et al., 1999). In contrast, the model of Dell et al. assumes a single lexical level while activation spreading is highly interactive (Dell et al., 1997). A similar debate between interactive accounts (MacDonald et al., 1994) and more modular models (Frazier, 1987; Friederici, 2002) has taken place in the area of sentence processing (see Friederici, 2002 for review).

In all models of word processing, repetition can be carried out along two independent routes. A "non-lexical route" maps perceived phonemes onto the response buffer and, subsequently, to articulation, while comprehension requires activation of semantic knowledge from auditory input. The conceptual representations may activate the word in a speaker's output lexicon, thus allowing for a second route for repetition.

A cognitive dual route model is also supported by the observations in the context of a recent case study from our laboratory (Bormann and Weiller, 2012). The subject in this study, BB, exhibited double dissociations between her ability to comprehend auditory words and to repeat these words. Occasionally, she would be able to repeat a word without comprehending its meaning (Examiner: "Please repeat the word 'hedgehog.' BB: 'Hedgehog, hedgehog,' I wonder what a hedgehog is."). On

Table 1 | Cognitive models.

SOUND

Wernicke	Freud	Morton	Dell	Levelt	Marslen-Wilson and Gaskell, Lambon Ralph
Word image	Word image	Buffer	Segmental phonology	Phonemes	Segmental phonology
Word concept	Word—object—association	Logogens	Lemmas	Lexemes	(hidden units) [no local lexicon]
				Lemmas	
Concepts	Sum of all associations	Semantic system	Semantic features	Concepts	Semantic features

THOUGHT

Cognitive models propose several levels of processing but all of them separate conceptual from linguistic representations. Models sometimes use different expressions for similar processes, and a level in one model may overlap with a level in another model. In older models (Wernicke, Freud), no phonemes are discerned.

other occasions, she would indicate comprehension of the word without being able to repeat it (Examiner: "Please repeat the word 'hedge.' BB: 'I know this one!' It's green, it may surround the garden, it may grow really high. But what was the word exactly?"). These examples provide evidence for a double dissociation between comprehension of a word's meaning and the ability to repeat words without comprehension. On the other hand, when available, both routes interact and contribute to word repetition (Jefferies et al., 2005).

The two routes assumed in cognitive models may be associated with the dorsal and ventral routes identified in our anatomic studies. The logogen model's non-lexical route may correspond to the dorsal route while comprehension of a word's or sentence's meaning may be associated with the ventral pathway. Our subject BB had severe deficits repeating non-words, so for repetition, she was mainly relying on the word's meaning. Semantic errors in repetition (Examiner: "Please repeat the word 'raw'." BB: "Was it 'meat'. I should say it was 'meat'") further support this conclusion. Within a cognitive model, her lesions would affect phonology to articulatory mapping and a predominant reliance upon a word's meaning during repetition.

While it is generally accepted that language as other cognitive functions emerges from context dependent interaction in distributed, segregated and overlapping networks (Damasio, 1989; Mesulam, 1990), this knowledge has not really spread to patients studies. With the advent of DTI tracking techniques the network architecture of functions is reflected anatomically. If the dorsal tract would be needed for repetition and the ventral tract for comprehension (Saur et al., 2008), it seems attractive to attribute isolated repetition problems to lesions of the dorsal tract and relate transcortical sensory aphasia, i.e., mainly comprehension deficits, to lesions of the ventral tract. However, attributing functions to tracks rather than to regions would just shift the focus but not solve the general problem of phrenology. In a network approach, the different cortical regions interact closely, and thus are not independent. Interruption of the network has an impact on the remaining ("intact") parts of the network. In other words, the functions of Broca's or Wernicke's area with an intact AF may not be the same as after that fascicle's destruction. This view does not contradict the assumption that a particular connection in a network on its own is necessary for e.g., repetition. However, the destruction of the interconnection may not result in a solitary repetition failure, but in a complete new phenomenological constellation, as the tract lesion affects the function of the regions it connects, and other regions in the remaining network, including in the other hemisphere, may become operational (Weiller et al., 1995). These assumptions make the complexity and fuzzy link of aphasic syndromes to focal lesions more understandable.

The difference may become clearer when comparing the method of symptom-lesion-mapping (e.g., Bates et al., 2003; Dronkers et al., 2004) with the "natural" occurrence of aphasic syndromes, i.e., compositions of symptoms. Across a series of 100 consecutive acute aphasics, comprehension problems (a score construed from tests examining: following verbal commands, word and sentence comprehension and results in the token test) using the Brunner–Munzel test map onto the EmC. Repetition problems (automated sequences as numbers or weekdays) map onto the AF (Kümmerer et al., 2010). Does this mean that comprehension is exclusively coded for or represented in the EmC? Rather not: in a network approach the observed comprehension deficit is produced by a lesion of the widespread "comprehension network." Why then do the infarcts of patients with comprehension deficit map onto the EmC, rather than let's say, Wernicke's area? The reason may be that the EmC holds a strategic position within the comprehension network. Within the distributed "comprehension network" (Friederici et al., 1999; Vigneau et al., 2006) all cortical regions in the temporal lobe (anterior and posterior MTG, fusiform gyrus) are connected with regions in the inferior frontal gyrus (BA 45 and BA 47) via a relatively small tract through the EmC (Saur et al., 2010). The EmC represents a kind of bottleneck of the network or in other words is the site where a small lesion can affect the entire network. Simultaneously it is situated in the middle of the middle cerebral artery (MCA) territory. Therefore, chances are highest that infarcts of the anterior or posterior MCA territory overlap in this small region, causing an interruption of the comprehension network.

This kind of lesion-mapping approach assumes that comprehension problems are independent from other potential deficits. However, only looking at comprehension problems is an "unnatural" setting. In most cases single symptoms should not be seen in isolation. There are only very few patients with solely a comprehension (or a repetition) problem. Symptoms of patients with aphasia cluster in syndromes that can be defined and reliably classified. This is due to the irrigation territories of the MCA and due to the organization of language in the brain. Can the syndrome of Wernicke's aphasia be explained by a lesion of the posterior part of the temporal lobe (e.g., Wernicke's area)? Comprehension problems are a hallmark of Wernicke's aphasia and indeed most people would agree that the temporal lobe does play a role in semantics. But Wernicke aphasia is characterized by more features, e.g., fluent speech with prominent semantic or phonemic paraphasias, thus also a "defect" in speech production. Wernicke's area participates in the ventral as well as the dorsal pathway and a lesion there will have to affect the functioning of both pathways. While semantic jargon may be due to comprehension problems, referring to the ventral pathway, phonemic jargon may be related to a lesion of the dorsal pathway. The sylvian parieto-temporal region (Area SPT) participates in sound-to-articulation mapping in the dorsal pathway (Hickok and Poeppel, 2000). As Wernicke himself put it: "the word images do not take appropriate control over the motor images" (Wernicke, 1874). Thus, Wernicke's aphasia represents a new phenomenological constellation, a syndrome, more than just comprehension problems. The occurrence of this syndrome is explained by the irrigation of the posterior temporal lobe by a branch of the MCA and its presentation is due to the lesion of the temporal lobe and the lesion of both the ventral and the dorsal pathway. For this interpretation we do not have to assume any form of diaschisis or reorganization.

Conduction aphasia due to a lesion of the AF is another problem. Destruction of the AF affects the dorsal pathway, but the arcuate fascicle does not in itself contain the representation for repetition, instead the pathway's functionality is needed also for repetition, and vice versa repetition may relay on the ventral pathway as well. Repetition should be altered and comprehension may be intact (via the ventral pathway) in most cases with AF lesions. Repetition of pseudowords also activates the ventral system (Saur et al., 2008), but this may merely reflect lexical search when trying to identify pseudowords. However, there is patient evidence that both routes contribute to repetition. The rare syndrome of deep dysphasia where subjects make semantic errors in single word repetition (e.g., repeating "crown" as "king") suggests a role of semantics in repetition. In addition, aphasic patients are better in repeating words they comprehend and are able to name in picture naming tasks in comparison to words which they do not understand (Jefferies et al., 2005). Semantic deficits may lead to mild repetition impairments because of the reduced support from meaning (Jefferies et al., 2005). Thus, repetition may be affected also outside conduction aphasia and lesions of the AF. Moreover, conduction aphasia is not restricted to repetition problems. What else would we have to expect with lesions of the AF, when referring to the dual loop system? Paraphasia through incorrect sensorimotor mapping in the dorsal pathway, "conduit d' approche" aiming to correct along the ventral pathway and also working memory deficits are often reported in patients with conduction aphasia. As in Wernicke's type aphasia, conduction aphasia is not to be reduced to repetition problems, and repetition not reduced to the AF.

The two pathways have different computational abilities, which are a prerequisite for various functions, depending on the modality. Lesions of a tract do affect the working of the entire network, resulting in a new phenomenological constellation, the syndrome is patient is presenting with. This syndrome is different from the loss of the supposed function, mediated by the pathway. Similarly, neglect and extinction can be differentiated by different effects of the lesions on the visuo spatial attention system (Umarova et al., 2011).

FROM SOUND TO CONCEPT AND BACK: THE DEVELOPMENT OF INNER SPEECH AND THE DUAL LOOP MODEL

So far, language in humans was discussed from an evolutionary view and in the perspective of cognitive models applied to patients with aphasia. However, the data do not yet explain how humans acquire the capacity for higher order thinking or abstraction. In this section, we want to argue that the development of "inner speech" in humans might be crucial for a simultaneous and close interaction between the two pathways, enabling the combination of phonological and abstract thought.

Jackendoff defined language as "essentially a mapping between sound and propositional or conceptual thought" (Jackendoff, 2009). Exactly how this mapping is achieved has been investigated in the context of psycholinguistic studies of language processing. Most cognitive models conceive of conceptual representations as independent from linguistic knowledge (**Table 1**) (Wernicke, 1906; Morton and Patterson, 1980; Freud, 1891; Dell et al., 1997; Gaskell and Marslen-Wilson, 1997; Levelt et al., 1999; Lambon Ralph et al., 2002). An aphasic speaker may be perfectly aware of the concept he or she is trying to name yet may be unable to access the word. Likewise, severely aphasic individuals may be unimpaired in tests of non-verbal reasoning, problem-solving and memory (Kertesz and Mccabe, 1975). There are several other examples of non-verbal thinking, such as face processing, mental rotation, spatial navigation, or tool use. Several contents of our mind do not map on lexical concepts and linguistic operations. In contrast, some aspects of language require less cognitive processing, e.g., the generation of automatic phonological sequences like counting or generating days of the week.

In most models, the process from sound to thought (and back) passes different levels of representation (**Table 1**). However, Vygotzky suggested that abstract thought processes and speech overlap, like two intersecting circles (Vygotzky, 1934). The overlapping part of thought and speech represents so-called "inner speech" or, as Vygotzky put it: "In their overlapping parts, thought and speech coincide to produce what is called verbal thought or inner speech, depending on the point of view." In inner speech, several aspects from purely sensorimotor to more abstract can be discerned. Especially in the early French literature, inner speech ("notion du mot") was seen purely as a means of (internal) sensorimotor mapping, or connecting phonological input and output properties, of course then related to working memory. Wernicke stated "The main task of the child that learns to speak is the imitation of the heard word" and this task, according to Wernicke, was performed by the direct (nowadays: dorsal) connection. "Only later the child is able to bind the word with a defined concept, long after the word has become a vast asset" (Wernicke, 1874).

But later on words would be spoken via the semantic route: "Soon after we have learned to speak a word, we lose the intention only to reproduce sounds and plan to utter a meaning" (Wernicke, 1874). "We have to assume that (then) the majority of speech impulses reach the word concepts from the remaining cerebral cortex" (Wernicke, 1906). Recently, Oppenheim and Dell showed that inner speech is more abstract on a phonological level, because covert segmental errors produced with tongue twisters were less similar to the target phoneme than overt slips of the tongue (Oppenheim and Dell, 2008). Therefore, internalization of speech is more than internalizing the production and anticipation of sound. It is combined with a concomitant increasing understanding what these phonological internal representations mean.

Only few patient studies exist investigating this topic. In a rhyming paradigm, where orthography alone was not sufficient to determine the sound of the rhyme, Geva et al. showed that in aphasic subjects, deficits of inner (= covert) speech, over and above deficits in overt speech production and working memory, were found following lesions to the left pars opercularis of the inferior frontal gyrus (BA 44) and the SMG (Geva et al., 2011). The authors concluded that for this aspect of inner speech, it was mainly the dorsal pathway that was affected. A recent paper shows that at birth, anterior and posterior language zones can be activated specifically but are not yet fully functionally connected (Perani et al., 2011), and that the interaction between the two regions becomes significantly synchronized around 7 years of age (Friederici et al., 2011). This finding was put in relation to the fact that the dorsal pathway has not yet fully matured (Brauer et al., 2011) and that children up to the age of seven are rather poor at comprehending syntactically complex sentences (Hahne et al., 2004; Dittmar et al., 2008; Dubois et al., 2008).

Also the full development of inner speech appears to occur around this age. According to Vygotsky, we do have conceptual awareness at birth, but no inner speech (Vygotzky, 1934). Young children start by accompanying their actions with speech, which evolves into "egocentric" or "private" speech while thinking aloud around the age of 4. In the process of a few years till the age of around 7 years, egocentric speech is replaced by inner speech (Vygotzky, 1934; Ehrich, 2006; Wiley, 2006). Behavioral studies in normally developing children and those with attention deficit hyperactivity disorder (ADHD) and autism support the importance of this process. The amount of internal speech in children correlates with performance (Winsler et al., 2000, 2003; Winsler and Naglieri, 2003; Ostad and Sorensen, 2007). Children with ADHD seem to have a delayed development of internalization (Berk and Potts, 1991), while on the other hand, in high performing children with autism, private speech remains relevant for performance (Winsler et al., 2007).

All these different aspects suggest that it is only after the internalization of speech and complete interaction of both pathways within the dual loop model that it is possible to simultaneously combine phonological and abstract thought simultaneously (**Figure 2**). This might also be a reason why complex grammar, for which simultaneous analysis of both time-dependent and time-independent processing is required, can only be mastered after the age of seven. Only then can we use language as a tool to represent abstract concepts (Deutscher, 2005), and, as Jackendoff puts it, use language as a "scaffolding that makes possible certain varieties of reasoning more complex than are available to non-linguistic organisms" (Jackendoff, 1997), permitting self-description and reflection, self-questioning and problem-solving, (Diaz and Berk, 1992; Barkley, 2001).

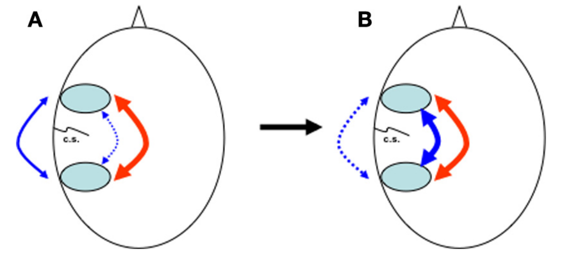


FIGURE 2 | Heads seen from above with pre- and post-rolandic areas around the central sulcus (c.s.). (A) At birth, dorsal anatomical connections (blue, dotted line) between pre- and post-rolandic areas are present but immature. In the first years, influenced by the continuous percept and anticipation of the consequence of movement and speech via feedback (blue, continuous line), internal connections and representations synchronise and mature. **(B)** Over time, external behaviour (blue, dotted line) is increasingly replaced by internal representations (blue, continuous line), while the interaction between dorsal (blue, continuous line) and ventral (red) pathways increases. For novel tasks, the external pathway can still be used.

It is suggested that during this crucial age of 4–7 years in children, not only in language, but in all modalities the emergence of internal representations takes place (Vygotzky, 1978; Diaz and Berk, 1992), "following the same general sequence of stages as the internalization of speech" and outer-directed behavior becomes turned on the self as a means to control one's own behavior (Barkley, 2001).

COMPARING THE DUAL LOOP MODEL IN LANGUAGE WITH OTHER MODALITIES: WHAT ARE THE ESSENTIAL CHARACTERISTICS OF THE VENTRAL AND DORSAL PATHWAYS?

There is no reason to suppose that the organization of the acoustic language system is different from other modalities. Current scientific evidence shows that a dual loop model, consisting of a dorsal and ventral pathway, can be found in different modalities, providing a scaffolding system for processing. In this dual loop model, hierarchy is not determined by one specific pathway, although specific functions may primarily or crucially involve one of both, but rather by an extension of this system in humans to regulatory, cytoarchitectonically more developed areas in prefrontal, temporal, and parietal neocortex (Weiller et al., 2011).

In **Table 2** we summarize some of the main studies reporting the involvement of a dorsal or a ventral pathway by different tasks. Also listed are a series of (partly unpublished) studies from our lab, in which we used DTI-based fiber tracking (Kreher et al., 2008) to connect seed regions in post- and prerolandic brain regions active during fMRI. In all examined modalities (language, motor, attention), we found ventral and dorsal connections along the EmC and the AF/SLF systems, respectively (**Figure 1**). Thus, functions ascribed to the dorsal and ventral pathway in

Table 2 | Summary of possible functions per modality processed along the two pathways from different studies.

Modality	Dorsal pathway	Ventral pathway
Vision	Spatial vision (a)	Object vision (b)
Acoustic	Sound localisation (c) Spatial working memory (c1) Integration of a target in a context (c2) "Temporarily buffering" the input (c3)	Sound Identification (d)
Language	"Mapping sound onto articulation" (e) Phonological loop (e1) Syntax (e2)	Semantic processing (f) Echoic or perceptual memory (f1) Recognition of perceptual incongruence (f2) Syntax (f3)
Motor	Control of actions "online" (g) Meaningless imitation (g1)	Motor imagination (h) Pantomime (h1)
Attention	Attention orientation (i)	Conscious perception of space (j)
Music	Recognition of structural incongruence (k) Tonal loop (k1)	Recognition of structural and perceptual incongruence (l)
Synthesis	Time-dependent Sequence execution (doing) Integration of forward and inverse models	Time-independent Meaning (understand what you're doing) Connection to world knowledge and concepts

[a] *The dorsal visual "action" pathway projects from early visual areas to the posterior parietal cortex and is engaged in visually guided actions as shown in experiments in golden hamsters (*Schneider, 1969*), in non-human primates (*Trevarthen, 1968; Mishkin et al., 1983; Desimone and Ungerleider, 1986; Bear et al., 2007*) and in humans (*Goodale and Milner, 1992; Wilson et al., 1993; Ungerleider and Haxby, 1994; Milner and Goodale, 1995; Rizzolatti and Matelli, 2003*).*

[b] *The ventral visual "perceptual" pathway, which projects from primary visual areas to the inferior temporal cortex, is crucial for object recognition (*Trevarthen, 1968; Mishkin et al., 1983; Desimone and Ungerleider, 1986; Gross, 1992; Wilson et al., 1993; Ungerleider and Haxby, 1994*).*

[c] *The caudal belt and parabelt regions interacting dorsally with the inferior parietal area are involved in sound localisation [c: (*Lewis and Van Essen, 2000*)]; while its interaction to dorsolateral prefrontal cortex is crucial for spatial working memory (c1) (*Rauschecker, 1995, 2011; Romanski et al., 1999a; Romanski and Goldman-Rakic, 2002; Rauschecker and Scott, 2009*) and for integration of a target in a context (c2) (*Rauschecker, 1995; Belin and Zatorre, 2000; Hickok and Poeppel, 2000, 2004; Scott and Wise, 2004*). The dorsal pathway subserves also the perception of the evolution over time of a sound in its spectral dynamics (c3) (*Rauschecker, 1995; Belin and Zatorre, 2000*).*

[d] *The identification of acoustic features of a sound involves the ventral pathways along temporal regions to ventrolateral prefrontal cortex in non-humans primates (*Romanski et al., 1999a,b; Belin and Zatorre, 2000; Rauschecker and Tian, 2000; Zatorre and Belin, 2001; Romanski and Goldman-Rakic, 2002; Tallal and Gaab, 2006*) and in humans (*Binder, 2000; Binder et al., 2000; Rauschecker and Tian, 2000; Thierry et al., 2003a; Ahveninen et al., 2006; Rauschecker and Scott, 2009; Leaver and Rauschecker, 2010*).*

[e] *Evidence of an integration of auditory sensory input and motor speech systems (e) along the dorsal pathway came from Geschwind (*Geschwind, 1965, 1967, 1972*), and from the dual loop model (*Hickok and Poeppel, 2000, 2004, 2007*), as well as from a feed-forward model (*Rauschecker, 2011*). Functional neuroimaging data support this (*Paulesu et al., 1993; Arnott et al., 2004; Hickok and Poeppel, 2004; Scott and Wise, 2004; Demonet et al., 2005*). Clear evidence comes from a DTI and fMRI study on repetition of pseudowords (*Saur et al., 2008*) and from intra-operative electrical stimulation (*Mandonnet et al., 2007*). The dorsal fronto-parieto-temporal pathway starting in the left inferior temporal occipital junction and progressing through the caudal part of the left superior temporal region and the inferior SMG to the left inferior frontal gyrus is also involved in monitoring speech at phonological level (e1) (*Paulesu et al., 1993; Demonet et al., 1994; Price, 1998; Pugh et al., 2000; Jobard et al., 2003; Demonet et al., 2005; Bernal and Ardila, 2009*) and also in the transformation from acoustic to phonetic information (*Binder et al., 2000*). Intra-operative electrical stimulation gives evidence of a dorsal phonological pathway, connecting the inferior frontal cortex (IFC)/ventral premotor cortex and the supramarginalis gyrus/postero-superior temporal cortex via cortico-cortical connections (*Duffau et al., 2003a,b*) and the arcuate fasciculus (*Duffau et al., 2002*). At least a dorsal pathway connecting temporo-parietal regions with Broca's area is involved in finite and phrase-structure grammar (e2) (*Friederici et al., 2006; Musso et al., 2009*) as well as in gender processing (*Vidorreta et al., 2011*).*

[f] *The ventral pathway along the anterior part of the left superior temporal sulcus is involved in intelligible speech (*Scott et al., 2000*), along the anterior part of the left STG in accessing semantic contents from spoken words (but not environmental sounds) (*Thierry et al., 2003b*). Combining DTI and fMRI method has shown that the EmC network is dominant for sentence comprehension (versus pseudo-sentences) (*Saur et al., 2008*). Intra-operative electrical stimulation also gives evidence of a ventral semantic pathway, connecting the IFC/dorsolateral prefrontal cortex and the posterior temporal regions via the inferior fronto-occipital fasciculus (*Duffau et al., 2005*) and via EmC but not UF (*Duffau et al., 2009*). Buchsbaum et al. found that auditory-verbal working memory depends on a ventral "what" pathway when initial retrieval is based on an episodic or perceptual code (f1) (*Buchsbaum et al., 2005*). This form of "perceptual" memory is labeled echoic memory (*Watkins and Watkins, 1980; Cowan, 1984; Penney, 1989*). Musso et al. found (f2) that perceptual action violation within a sentence relates to an insulo-temporal interaction along left EmC (*Musso et al., 2009*). Grammar processing (f3) involves ventral fronto-temporo-parietal interaction for an artificial finite-state grammar task (*Friederici et al., 2006*) and for recognition of long term dependencies in real language (*Musso et al., 2009*).*

[g] *The major functional role of the dorsal pathway between the visual area and the superior parietal lobule is the control of actions "online" (g) (*Rizzolatti and Matelli, 2003*). The dorsal interaction between parietal and frontal (the pars opercularis of the inferior frontal gyrus) is related to motor planning (*Jeannerod, 1994*) and execution (*Stephan et al., 1995; Gerardin et al., 2000*). Imitation of meaningless action exclusively involves the dorsal SLF 2-3 pathway (g1) (*Vry et al., 2012*).*

(Continued)

Continued

[h] *Areas activated by movement imagination are more anterior and posterior than execution alone (Decety et al., 1994; Stephan et al., 1995; Gerardin et al., 2000; Hanakawa et al., 2003, 2008) and are connected by ventral connections (Vry et al., 2012). This system is also involved in understanding the meaning of the movement when pantomiming object use (h1) (Vry et al. in preparation).*

[i] *Corbetta and Shulman (Corbetta and Shulman, 2002)*

[j] *(Umarova et al., 2010)*

[k] *Integration of linguistic and musical elements within structural representations involves the same amplitude of the P600, a centroparietal component that, therefore, could be related to dorsal pathway (Patel, 2008). Musso et al. showed that a parieto-frontal along the SLF 2-3 pathway is involved in the detection of structural incongruence (a chord out of key) (Musso et al., 2009). Schulze et al. showed a specific involvement of pars opercularis and parieto-temporal activation (and, therefore probably related to the dorsal pathway) for tonal working memory (k1) (Schulze et al., 2011).*

[l] *In Musso et al. the ventral parieto-frontal interaction is required for recognition of structural as well as perceptual violations (a chord out of tune) (Musso et al., 2009).*

each modality may differ, but only in their modality-specific aspect.

A VENTRAL PATHWAY FOR TIME-INDEPENDENT PROCESSING

The ventral pathway was first described as the "what" pathway in the visual system of chimpanzees (Mishkin et al., 1983) and humans (Ungerleider and Haxby, 1994). Later, this function was assigned to the acoustic system as well. A ventral pathway, connecting the anterior belt and parabelt with the anterior temporal regions underlies auditory object identification (Romanski et al., 1999a,b; Rauschecker and Scott, 2009), with the anterior part of the left supratemporal gyrus (STG) involved in accessing semantic contents from spoken words vs. environmental sounds (Thierry et al., 2003b) as well from intelligible speech (Scott et al., 2000), while the inferior frontal gyrus (pars orbitalis) is necessary for sentence comprehension (as compared to pseudo-sentences) (Saur et al., 2008). Confirmation of the existence of a ventral semantic pathway, connecting the inferior frontal gyrus/dorsolateral prefrontal gyrus (IFG/DLPFG) and the posterior temporal regions via the inferior fronto-occipital fasciculus and EmC, but not the UF, comes from intra-operative electrical stimulation (Duffau et al., 2005, 2009). These data indicate that the antero-ventral speech processing pathway is crucial for mapping acoustic-phonemic cues onto lexical representations (Scott and Johnsrude, 2003).

Is thus "meaning" the more general function of this pathway? Several kinds of evidence show that the ventral pathway is involved in the identification of adjacent as well as non-adjacent syntactic relations of the perceived linguistic elements (Friederici et al., 2006; Musso et al., 2009) and of tonal dependencies (Musso et al., 2009). The ventral route is thus involved in the identification of structural relations independent of the modality and of the time of occurrence of each element.

A ventral pathway through the EmC seems to play an important and similar role also in other modalities. In the attentional system, a ventral tract, connecting the parietal and temporal lobe with the anterior insula and the ventrolateral prefrontal cortex, was interpreted as being critical for the integration of the perception of space for an intended action and for the correct estimation of the relevance of stimuli to the self (Umarova et al., 2010). In the motor system, areas in posterior parietal and prefrontal cortices involved in imagery of movements (Vry et al., 2012) as well in pantomiming object use (Vry et al., in preparation) are connected via the ventral route, putting the ventral tract in relation to symbolic acts and cognition.

We speculate that a more general function of the ventral system could be the extraction, relation and implementation of an invariant set of properties of the perceived elements (in a limited number of given possibilities, i.e., as "a priori" categories), which are related to semantic memory and meaning. Processing along the ventral pathway is, therefore not dependent on the temporal or spatial sequence of elements, rather it is optimized to test a limited number of possible combinations in order to extract meaning (Weiller et al., 2011).

A DORSAL PATHWAY FOR TIME-DEPENDENT PROCESSING

The dorsal route was initially labeled the "where" pathway as it is found to be involved in processing spatial relations between visual or acoustic perceived objects as well as between oneself and external objects to identify visual motion and is used for the visual control of action (see **Table 2, i–k**). The term "how" pathway for the dorsal pathways was introduced later and relates to sensory-motor integration function (Kravitz et al., 2011). Individuals with brain damage of the dorsal visual pathway affecting the posterior and the superior parietal cortex suffer from optic ataxia. In this condition, a deficit of the visuomotor system, the size, shape, and color and even the location (the "where") of the object remain intact, but the ability to identify object arrangements and to perform "goal-directed actions to visual targets" with any sort of precision and accuracy is disturbed (Perenin and Vighetto, 1988; Goodale et al., 1991; Milner et al., 2003). The auditory dorsal pathway is predominantly related to speech production (e) and was demonstrated to be involved in non-word repetition, thereby providing a phonology-to-articulation interface for correct speech (see **Table 2**, e1). Because the representation in semantic memory is lacking, the sensory percept of pseudowords has to be mapped on the motor representations for repetition.

However, the function of the acoustic dorsal pathway seems not to be limited to "mapping sound onto articulation" (**Table 2**), but rather serves to integrate linguistic or musical syntactic elements in a context (Patel, 2008; Musso et al., 2009), even to process phrase-structure grammar (Friederici et al., 2006), or to integrate and maintain the perceived auditory signals within context over temporal and spatial evolution (see **Table 2, c2,3, e1**).

Thus, the acoustic dorsal pathway would not be limited to "where" or "how" functions, rather its more general role, independent from the modality, is the capacity to analyze the sequence of segments, either in time or in space, as well as fast online integration between sensory event information and "internal models or emulators (Rauschecker and Scott, 2009). Spatial transformation as well as sensorimotor integration may be examples of adaptations used by forward models (predictors) and inverse models (controllers) (Rauschecker and Scott, 2009).

Through exercise and experience, moulds are developed, which can be called on quickly. By learning a movement like a signature, parameters for this movement are stored and can be accessed by another extremity on demand (Rijntjes et al., 1999). Thus, in the dorsal pathway, in contrast to the ventral pathway, stable connotations like blueprints can be developed in an infinite number of possibilities, the only constraint being the physical and computational limitations of movement themselves.

The mechanism for "online" analysis of sequences may be seen as a function of the dorsal system in other modalities as well. A dorsal network was recently described in the attentional network, where dorsal pathways along the superior longitudinal fascicle/AF system connected the parietal and temporal lobe with the premotor cortices (BA 6, 44, 8) and was interpreted to convey information needed for spatial stimulus orientation or processing of peri-personal space (**Tables 1, 2**) (Umarova et al., 2010). A similar pattern seems to exist in the motor system: during simple, repetitive, externally cued active or passive movements, a dorsal route connects the parietal cortex with premotor areas for sensorimotor mapping, motor control based on internal predictive models and sensory feedback (Wolpert et al., 1995; Wolpert and Miall, 1996; Grush, 2004; Shadmehr and Krakauer, 2008; Vry et al., 2012).

It is not clarified whether or not working memory predominantly uses the dorsal pathway. It is generally acknowledge that the frontal areas involved in working memory, as pars opercularis or precentral cortex, are mainly connected via the dorsal fiber system with parietal and temporal lobe (Paulesu et al., 1993; Wager and Smith, 2003). These tracts may be involved in short-term retention of the phonological input and lesions to the inferior parietal areas in the left hemisphere usually cause verbal short-term memory impairments (Vallar et al., 1997). In a recent fMRI study Buchsbaum et al. distinguish between perceptually based ("echoic") memory, which relates to the ventral pathway, from phonological-articulatory memory, which was confirmed to be a predominant dorsal task (Buchsbaum et al., 2005).

LANGUAGE: HOW UNIQUE IS IT?

In summary, after internalization of speech and movement, a dual loop system, consisting of efficiently and flexibly interacting dorsal and ventral pathways, extending to a highly developed gradient along pre- and post-rolandic regions, seems to be basis of similar functions in all modalities, including language. We started with the assumption that language is unique in humans, and there are indeed anatomical reasons to support this notion. However, since there are so many anatomical and functional similarities with other modalities, it seems justified to ask the question: to what extent do also other modalities have unique properties in humans?

Should we expect that also human motor processing and attention is different from lower primates? Are humans better in understanding the meaning of movement than animals? Can a chimpanzee understand the meaning if a human experimenter pantomimes peeling a banana? Humans may have a thicker AF, which may allow them to speak easily and a lot. Are they therefore better in motor skills as well, can they therefore perform a perfect serve in tennis, slalom on a steep skiing slope or perform a complex piano play "by heart"? If language, through inner speech, is a prerequisite for the wealth of our inner world, is there an equivalent of inner speech in other modalities, and what is it like?

Again, the development of the dorsal and ventral pathway, in the frontal lobe converging on Broca's area, could be the crucial anatomical feature. Numerous studies have shown an involvement of area 45 when processing hierarchical structures not only in the language modality (Musso et al., 2003), but also in movement (Binkofski and Buccino, 2004; Fiebach and Schubotz, 2006; Tettamanti and Weniger, 2006), music (Tettamanti and Weniger, 2006; Musso et al., 2009), and in the homologue area of the right hemisphere, for attention (Umarova et al., 2010). If so, it is an unresolved question whether the differentiation in dorsal and ventral connections in humans along an anterior-posterior gradient, especially in Broca's area, evolved primarily for language and that other modalities were able to use these highly differentiated functional structures, or that a parallel development enabled all modalities, including language, to attain human specific attributes.

REFERENCES

Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., Levanen, S., Lin, F. H., Sams, M., Shinn-Cunningham, B. G., Witzel, T., and Belliveau, J. W. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14608–14613.

Almeida, J., Mahon, B. Z., and Caramazza, A. (2010). The role of the dorsal visual processing stream in tool identification. *Psychol. Sci.* 21, 772–778.

Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H. B., and Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412, 319–341.

Anderson, B., Southern, B. D., and Powers, R. E. (1999). Anatomic asymmetries of the posterior superior temporal lobes: a postmortem study. *Neuropsychiatry Neuropsychol. Behav. Neurol.* 12, 247–254.

Anwander, A., Tittgemeyer, M., Von Cramon, D. Y., Friederici, A. D., and Knosche, T. R. (2007). Connectivity-based parcellation of Broca's area. *Cereb. Cortex* 17, 816–825.

Arnott, S. R., Binns, M. A., Grady, C. L., and Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage* 22, 401–408.

Badre, D., and D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat. Rev. Neurosci.* 10, 659–669.

Bamiou, D. E., Musiek, F. E., and Luxon, L. M. (2003). The insula (Island of Reil) and its role in auditory processing. Literature review. *Brain Res. Brain Res. Rev.* 42, 143–154.

Bamiou, D. E., Musiek, F. E., Stow, I., Stevens, J., Cipolotti, L., Brown, M. M., and Luxon, L. M. (2006). Auditory temporal processing deficits in patients with insular stroke. *Neurology* 67, 614–619.

Barkley, R. A. (2001). The executive functions and self-regulation: an evolutionary neuropsychological perspective. *Neuropsychol. Rev.* 11, 1–29.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., and Dronkers, N. F. (2003). Voxel-based lesion-symptom mapping. *Nat. Neurosci.* 6, 448–450.

Bear, M. F., Connors, B. W., and Paradiso, M. A. (2007). *Neuroscience: Exploring the brain*. Baltimore, MD: Lippincott Williams and Wilkins.

Belin, P., Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nat. Neurosci.* 3, 965–966.

Berk, L. E., and Potts, M. K. (1991). Development and functional significance of private speech among attention-deficit hyperactivity disordered and normal boys. *J. Abnorm. Child Psychol.* 19, 357–377.

Berke, J. J. (1960). The Claustrum, the external capsule and the extreme capsule of *Macaca mulatta*. *J. Comp. Neurol.* 115, 297–331.

Bernal, B., and Ardila, A. (2009). The role of the arcuate fasciculus in conduction aphasia. *Brain* 132, 2309–2316.

Binder, J. (2000). The new neuroanatomy of speech perception. *Brain* 123(Pt 12), 2371–2372.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528.

Binkofski, F., and Buccino, G. (2004). Motor functions of the Broca's region. *Brain Lang.* 89, 362–369.

Bormann, T., and Weiller, C. (2012). "Are there lexicons?" A study of lexical and semantic processing in word-meaning deafness suggests "yes." *Cortex* 48, 294–307.

Brauer, J., Anwander, A., and Friederici, A. D. (2011). Neuroanatomical prerequisites for language functions in the maturing brain. *Cereb. Cortex* 21, 459–466.

Buchel, C., Raedler, T., Sommer, M., Sach, M., Weiller, C., and Koch, M. A. (2004). White matter asymmetry in the human brain: a diffusion tensor MRI study. *Cereb. Cortex* 14, 945–951.

Buchsbaum, B. R., Olsen, R. K., Koch, P., and Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48, 687–697.

Bucy, P. C., and Kluver, H. (1955). An anatomical investigation of the temporal lobe in the monkey (*Macaca mulatta*). *J. Comp. Neurol.* 103, 151–251.

Butti, C., and Hof, P. R. (2010). The insular cortex: a comparative perspective. *Brain Struct. Funct.* 214, 477–493.

Buxhoeveden, D. P., and Casanova, M. F. (2002). The minicolumn and evolution of the brain. *Brain Behav. Evol.* 60, 125–151.

Buxhoeveden, D. P., Switala, A. E., Litaker, M., Roy, E., and Casanova, M. F. (2001a). Lateralization of minicolumns in human planum temporale is absent in nonhuman primate cortex. *Brain Behav. Evol.* 57, 349–358.

Buxhoeveden, D. P., Switala, A. E., Roy, E., Litaker, M., and Casanova, M. F. (2001b). Morphological differences between minicolumns in human and nonhuman primate cortex. *Am. J. Phys. Anthropol.* 115, 361–371.

Cantalupo, C., and Hopkins, W. D. (2001). Asymmetric Broca's area in great apes. *Nature* 414, 505.

Casanova, M. F., Trippe, J., Tilquist, C. R., and Switala, A. E. (2010). Dolphin insula reflects minicolumnar organization of mammalian isocortex. *Transl. Neurosci.* 1, 37–42.

Casanova, M. F., Van Kooten, I. A., Switala, A. E., Van Engeland, H., Heinsen, H., Steinbusch, H. W., Hof, P. R., Trippe, J., Stone, J., and Schmitz, C. (2006). Minicolumnar abnormalities in autism. *Acta Neuropathol.* 112, 287–303.

Caspers, S., Eickhoff, S. B., Rick, T., Von Kapri, A., Kuhlen, T., Huang, R., Shah, N. J., and Zilles, K. (2011). Probabilistic fibre tract analysis of cytoarchitectonically defined human inferior parietal lobule areas reveals similarities to macaques. *Neuroimage* 58, 362–380.

Catani, M., Allin, M. P., Husain, M., Pugliese, L., Mesulam, M. M., Murray, R. M., and Jones, D. K. (2007). Symmetries in human brain language pathways correlate with verbal recall. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17163–17168.

Catani, M., Howard, R. J., Pajevic, S., and Jones, D. K. (2002). Virtual *in vivo* interactive dissection of white matter fasciculi in the human brain. *Neuroimage* 17, 77–94.

Catani, M., Jones, D. K., and Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.

Cherniak, C. (1994). Component placement optimization in the brain. *J. Neurosci.* 14, 2418–2427.

Clarke, S., Bellmann Thiran, A., Maeder, P., Adriani, M., Vernet, O., Regli, L., Cuisenaire, O., and Thiran, J. P. (2002). What and where in human audition: selective deficits following focal hemispheric lesions. *Exp. Brain Res.* 147, 8–15.

Clarke, S., Bellmann, A., Meuli, R. A., Assal, G., and Steck, A. J. (2000). Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. *Neuropsychologia* 38, 797–807.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.

Corbetta, M., Kincade, M. J., Lewis, C., Snyder, A. Z., and Sapir, A. (2005). Neural basis and recovery of spatial attention deficits in spatial neglect. *Nat. Neurosci.* 8, 1603–1610.

Cowan, N. (1984). On short and long auditory stores. *Psychol. Bull.* 96, 341–370.

Creem, S. H., and Proffitt, D. R. (2001). Grasping objects by their handles: a necessary interaction between cognition and action. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 218–228.

Crick, F. C., and Koch, C. (2005). What is the function of the claustrum? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1271–1279.

Damasio, A. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput.* 1, 123–132.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., Mazziotta, J. C., and Fazio, F. (1994). Mapping motor representations with positron emission tomography. *Nature* 371, 600–602.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., and Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.* 104, 801–838.

Demonet, J. F., Price, C., Wise, R., and Frackowiak, R. S. (1994). A PET study of cognitive strategies in normal subjects during language tasks. Influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain* 117(Pt 4), 671–682.

Demonet, J. F., Thierry, G., and Cardebat, D. (2005). Renewal of the neurophysiology of language: functional neuroimaging. *Physiol. Rev.* 85, 49–95.

Desimone, R., and Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *J. Comp. Neurol.* 248, 164–189.

Deutscher, G. (2005). *The Unfolding of Language: The Evolution of Mankind's Greatest Invention.* London, UK: Random House.

Diaz, R. M., and Berk, L. E. (1992). *Private Speech: From Social Interaction to Self-Regulation.* Mahwah, NJ: Erlbaum.

Dittmar, M., Abbot-Smith, K., Lieven, E., Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Dev.* 79, 1152–1167.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D. J. R., Redfern, B. B., and Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition* 92, 145–177.

Dubois, J., Dehaene-Lambertz, G., Perrin, M., Mangin, J. F., Cointepas, Y., Duchesnay, E., Le Bihan, D., and Hertz-Pannier, L. (2008). Asynchrony of the early maturation of white matter bundles in healthy infants: quantitative landmarks revealed noninvasively by diffusion tensor imaging. *Hum. Brain Mapp.* 29, 14–27.

Duffau, H. (2012). The "frontal syndrome" revisited: lessons from electrostimulation mapping studies. *Cortex* 48, 120–131.

Duffau, H., Capelle, L., Denvil, D., Gatignol, P., Sichez, N., Lopes, M., Sichez, J. P., and Van Effenterre, R. (2003a). The role of dominant premotor cortex in language: a study using intraoperative functional mapping in awake patients. *Neuroimage* 20, 1903–1914.

Duffau, H., Gatignol, P., Denvil, D., Lopes, M., and Capelle, L. (2003b). The articulatory loop: study of the subcortical connectivity by electrostimulation. *Neuroreport* 14, 2005–2008.

Duffau, H., Capelle, L., Sichez, N., Denvil, D., Lopes, M., Sichez, J. P., Bitar, A., and Fohanno, D. (2002). Intraoperative mapping of the subcortical language pathways using direct stimulations. An anatomo-functional study. *Brain* 125, 199–214.

Duffau, H., Gatignol, P., Mandonnet, E., Peruzzi, P., Tzourio-Mazoyer, N., and Capelle, L. (2005). New insights into the anatomo-functional connectivity of the semantic system: a study using cortico-subcortical electrostimulations. *Brain* 128, 797–810.

Duffau, H., Gatignol, P., Moritz-Gasser, S., and Mandonnet, E. (2009). Is the left uncinate fasciculus essential for language? A cerebral stimulation study. *J. Neurol.* 256, 382–389.

Edelstein, L. R., and Denaro, F. J. (2004). The claustrum: a historical review of its anatomy, physiology, cytochemistry and functional significance. *Cell. Mol. Biol. (Noisy-le-grand)* 50, 675–702.

Ehrich, J. F. (2006). Vygotskian inner speech and the reading process. *Aust. J. Educ. Dev. Psychol.* 6, 12–25.

Fiebach, C. J., and Schubotz, R. I. (2006). Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains? *Cortex* 42, 499–502.

Frazier, L. (1987). "Sentence processing: a tutorial review," in *Attention and Performance, XII.* ed M. Coltheart (Hillsdale, NJ: Erlbaum).

Freud, S. (1891). *Zur Auffassung der Aphasien. Eine kritische Studie,* 2nd Edn. Leipzig, Wien: Franz Deuticke.

Frey, S., Campbell, J. S., Pike, G. B., and Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.* 28, 11435–11444.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84.

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends Cogn. Sci.* 13, 175–181.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., and Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2458–2463.

Friederici, A. D., Brauer, J., and Lohmann, G. (2011). Maturation of the language network: from inter—to intrahemispheric connectivities. *PLoS ONE* 6:e20726. doi: 10.1371/journal.pone.0020726

Friederici, A. D., Von Cramon, D. Y., and Kotz, S. A. (1999). Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain* 122(Pt 6), 1033–1047.

Galuske, R. A., Schlote, W., Bratzke, H., and Singer, W. (2000). Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science* 289, 1946–1949.

Gaskell, M. G., and Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Lang. Cogn. Proc.* 12, 613–656.

Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.

Gerardin, E., Sirigu, A., Lehericy, S., Poline, J. B., Gaymard, B., Marsault, C., Agid, Y., and Le Bihan, D. (2000). Partially overlapping neural networks for real and imagined hand movements. *Cereb. Cortex* 10, 1093–1104.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. *Brain* 88, 237-294, 585–644.

Geschwind, N. (1967). Wernickes contribution to the study of aphasia. *Cortex* 3, 449–463.

Geschwind, N. (1972). Language and the brain. *Sci. Am.* 226, 76–83.

Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J. C., and Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain* 134, 3071–3082.

Gloor, P. (1997). *The Temporal Lobe and Limbic System.* New York, NY: Oxford University Press.

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25.

Goodale, M. A., Milner, A. D., Jakobson, L. S., and Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature* 349, 154–156.

Gross, C. G. (1992). Representation of visual stimuli in inferior temporal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 335, 3–10.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–96; discussion 396–442.

Hagmann, P., Cammoun, L., Martuzzi, R., Maeder, P., Clarke, S., Thiran, J. P., and Meuli, R. (2006). Hand preference and sex shape the architecture of language networks. *Hum. Brain Mapp.* 27, 828–835.

Hahne, A., Eckstein, K., and Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *J. Cogn. Neurosci.* 16, 1302–1318.

Hanakawa, T., Dimyan, M. A., and Hallett, M. (2008). Motor planning, imagery, and execution in the distributed motor network: a time-course study with functional MRI. *Cereb. Cortex* 18, 2775–2788.

Hanakawa, T., Immisch, I., Toma, K., Dimyan, M. A., Van Gelderen, P., and Hallett, M. (2003). Functional properties of brain areas associated with motor execution and imagery. *J. Neurophysiol.* 89, 989–1002.

Hauser, M. D., Newport, E. L., and Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: statistical learning in cotton-top tamarins. *Cognition* 78, B53–B64.

Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138.

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.

Hoeft, F., Barnea-Goraly, N., Haas, B. W., Golarai, G., Ng, D., Mills, D., Korenberg, J., Bellugi, U., Galaburda, A., and Reiss, A. L. (2007). More is not always better: increased fractional anisotropy of superior longitudinal fasciculus associated with poor visuospatial abilities in Williams syndrome. *J. Neurosci.* 27, 11960–11965.

Horton, J. C., and Hedley-Whyte, E. T. (1984). Mapping of cytochrome oxidase patches and ocular dominance columns in human visual cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 304, 255–272.

Jackendoff, R. (1997). *The Architecture of the Language Faculty.* Cambridge, MA: The MIT Press.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Percept.* 26, 195–204.

Jeannerod, M. (1994). The representing brain: neural correlates of motor intention and imagery. *Behav. Brain Sci.* 17, 187–245.

Jefferies, E., Jones, R. W., Bateman, D., and Ralph, M. A. (2005). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Cogn. Neuropsychol.* 22, 183–212.

Jobard, G., Crivello, F., and Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *Neuroimage* 20, 693–712.

Kastner, S., and Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23, 315–341.

Kertesz, A., and Mccabe, P. (1975). Intelligence and aphasia: performance of aphasics on Raven's coloured progressive matrices (RCPM). *Brain Lang.* 2, 387–395.

Kravitz, D. J., Saleem, K. S., Baker, C. I., and Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nat. Rev. Neurosci.* 12, 217–230.

Kreher, B. W., Schnell, S., Mader, I., Il'yasov, K. A., Hennig, J., Kiselev, V. G., and Saur, D. (2008). Connecting and merging fibres: pathway extraction by combining probability maps. *Neuroimage* 43, 81–89.

Krubitzer, L. (1995). The organization of neocortex in mammals: are species differences really so different? *Trends Neurosci.* 18, 408–417.

Kümmerer, D., Kellmeyer, P., Glauche, V., Mader, I., Kloppel, S., Suchan, J., Karnath, H. O., Weiller, C., and Saur, D., (2010). "Damage of white matter fiber tracts in acute aphasic patients-a voxelwise lesionbehavior mapping study," in *11th international Science of Aphasia Conference,* (Potsdam, Germany). Available at: http://www.soa-online.com/programandlowbar;assets/final%20program%20abstracts.pdf

Lambon Ralph, M. A., Moriarty, L., and Sage, K. (2002). Anomia is simply a reflection of semantic and phonological impairments: evidence from a case-series study. *Aphasiology* 16, 56–82.

Leaver, A. M., and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612.

Leavitt, V. M., Molholm, S., Gomez-Ramirez, M., and Foxe, J. J. (2011). "What" and "where" in auditory sensory processing: a high-density electrical mapping study of distinct neural processes underlying sound object recognition and sound localization. *Front. Integr. Neurosci.* 5:23. doi: 10.3389/fnint.2011.00023

Lebel, C., Walker, L., Leemans, A., Phillips, L., and Beaulieu, C. (2008). Microstructural maturation of the human brain from childhood to adulthood. *Neuroimage* 40, 1044–1055.

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38; discussion 38–75.

Lewis, J. W., and Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 428, 112–137.

Livingstone, M. S., and Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *J. Neurosci.* 4, 309–356.

Ludwig, E., and Klinger, J. (1956). *Atlas Cerebri Humani.* Basel, New York, NY: Karger.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution [corrected]. *Psychol. Rev.* 101, 676–703.

Mahon, B. Z., Milleville, S. C., Negri, G. A., Rumiati, R. I., Caramazza, A., and Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron* 55, 507–520.

Makris, N., and Pandya, D. N. (2009). The extreme capsule in humans and rethinking of the language circuitry. *Brain Struct. Funct.* 213, 343–358.

Makris, N., Meyer, J. W., Bates, J. F., Yeterian, E. H., Kennedy, D. N., and Caviness, V. S. (1999). MRI-Based topographic parcellation of human cerebral white matter and nuclei II. Rationale and applications with systematics of cerebral connectivity. *Neuroimage* 9, 18–45.

Mandonnet, E., Nouet, A., Gatignol, P., Capelle, L., and Duffau, H. (2007). Does the left inferior longitudinal fasciculus play a role in language? A brain stimulation study. *Brain* 130, 623–629.

Manger, P., Sum, M., Szymanski, M., Ridgway, S. H., and Krubitzer, L. (1998). Modular subdivisions of dolphin insular cortex: does evolutionary history repeat itself? *J. Cogn. Neurosci.* 10, 153–166.

Mathur, B. N., Caprioli, R. M., and Deutch, A. Y. (2009). Proteomic analysis illuminates a novel structural definition of the claustrum and insula. *Cereb. Cortex* 19, 2372–2379.

Mesulam, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann. Neurol.* 28, 597–613.

Mesulam, M. M. (1998). From sensation to cognition. *Brain* 121(Pt 6), 1013–1052.

Meynert, T. (1866). "Ein Fall von Sprachstörung, anatomisch begründet," in *XII. Band der Zeitschrift der*, eds C. Braun, A. Duschek, and L. Schläger (Wien: K.u.K. Gesellschaft der Ärzte in Wien), 152–189.

Milner, A. D., and Goodale, M. A. (1995). *The Visual Brain in Action.* Oxford, Oxford University Press.

Milner, A. D., Dijkerman, H. C., Mcintosh, R. D., Rossetti, Y., and Pisella, L. (2003). Delayed reaching and grasping in patients with optic ataxia. *Prog. Brain Res.* 142, 225–242.

Mishkin, M., Ungerleider, L., and Macko, K. A. (1983). Object vision and spatial vision: two visual pathways. *Trends Neurosci.* 6, 414–417.

Morton, J., Patterson, K. E. (1980). "A new attempt at an interpretation, or, an attempt at a new interpretation," in *Deep Dyslexia,* eds M. Coltheart, K. E. P., and J. C. Marshall (London, UK: Routledge and Kegan Paul).

Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.* 20, 408–434.

Mountcastle, V. B. (1978). "An organizing principle for cerebral function: the unit module and the distributed system," in *The Mindful Brain: Cortical Organization and the Group-Selective Theory,* eds G. M. Edelman and V. B. Mountcastle (Cambridge, MA: MIT Press).

Musso, M., Glauche, V., Horn, A., and Weiller, C. (2009). The signature of human syntactic architecture. *Neuroimage* 47, S118.

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Buchel, C., and Weiller, C. (2003). Broca's area and the language instinct. *Nat. Neurosci.* 6, 774–781.

Northcutt, R. G., and Kaas, J. H. (1995). The emergence and evolution of mammalian neocortex. *Trends Neurosci.* 18, 373–379.

Nucifora, P. G., Verma, R., Melhem, E. R., Gur, R. E. G. U. R. R. C. (2005). Leftward asymmetry in relative fiber density of the arcuate fasciculus. *Neuroreport* 16, 791–794.

Oechslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., and Jancke, L. (2009). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. *Front. Hum. Neurosci.* 3:76. doi: 10.3389/neuro.09.076.2009

Oppenheim, G. M., and Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* 106, 528–537.

Ostad, S. A., and Sorensen, P. M. (2007). Private speech and strategy-use patterns: bidirectional comparisons of children with and without mathematical difficulties in a developmental perspective. *J. Learn. Disabil.* 40, 2–14.

Parker, G. J., Luzzi, S., Alexander, D. C., Wheeler-Kingshott, C. A., Ciccarelli, O., and Lambon Ralph, M. A. (2005). Lateralization of ventral and dorsal auditory-language pathways in the human brain. *Neuroimage* 24, 656–666.

Patel, A. D. (2008). *Music, Language, and the Brain.* Oxford, UK: Oxford University Press.

Paulesu, E., Frith, C. D., and Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature* 362, 342–345.

Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, N. A., De Giovanni, U., Sensolo, S., and Fazio, F. (2003). A functional-anatomical model for lipreading. *J. Neurophysiol.* 90, 2005–2013.

Paus, T., Zijdenbos, A., Worsley, K., Collins, D. L., Blumenthal, J., Giedd, J. N., Rapoport, J. L., and Evans, A. C. (1999). Structural maturation of neural pathways in children and adolescents: *in vivo* study. *Science* 283, 1908–1911.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Mem. Cognit.* 17, 398–422.

Pepperberg, I. M., and Shive, H. R. (2001). Simultaneous development of vocal and physical object combinations by a Grey parrot (Psittacus erithacus): bottle caps, lids, and labels. *J. Comp. Psychol.* 115, 376–384.

Perani, D., Saccuman, M. C., Scifo, P., Awander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., and Friederici, A. D. (2011). Neural language networks at birth. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16056–16061.

Perenin, M. T., and Vighetto, A. (1988). Optic ataxia: a specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. *Brain* 111(Pt 3), 643–674.

Peters, S. U., Kaufmann, W. E., Bacino, C. A., Anderson, A. W., Adapa, P., Chu, Z., Yallampalli, R., Traipe, E., Hunter, J. V., and Wilde, E. A. (2011). Alterations in white matter pathways in Angelman syndrome. *Dev. Med. Child Neurol.* 53, 361–367.

Petrides, M., and Pandya, D. N. (1988). Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J. Comp. Neurol.* 273, 52–66.

Petrides, M., and Pandya, D. N. (2007). Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *J. Neurosci.* 27, 11573–11586.

Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7:e1000170. doi: 10.1371/journal.pbio.1000170

Powell, H. W., Parker, G. J., Alexander, D. C., Symms, M. R., Boulby, P. A., Wheeler-Kingshott, C. A., Barker, G. J., Noppeney, U., Koepp, M. J., and Duncan, J. S. (2006). Hemispheric asymmetries in language-related pathways: a combined functional MRI and tractography study. *Neuroimage* 32, 388–399.

Price, C. J. (1998). The functional anatomy of word comprehension and production. *Trends Cogn. Sci.* 2, 281–288.

Pugh, K. R., Mencl, W. E., Jenner, A. R., Katz, L., Frost, S. J., Lee, J. R., Shaywitz, S. E., and Shaywitz, B. A. (2000). Functional neuroimaging studies of reading and reading disability (developmental dyslexia). *Ment. Retard. Dev. Disabil. Res. Rev.* 6, 207–213.

Purves, D., Riddle, D. R., and Lamantia, A. S. (1992). Iterated patterns of brain circuitry (or how the cortex gets its spots). *Trends Neurosci.* 15, 362–368.

Rakic, P. (1995). A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci.* 18, 383–388.

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.

Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11800–11806.

Rauschecker, J. P. (1995). Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends Neurosci.* 18, 36–43.

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear. Res.* 271, 16–25.

Rijntjes, M., Dettmers, C., Buchel, C., Kiebel, S., Frackowiak, R. S., and Weiller, C. (1999). A blueprint for movement: functional and anatomical representations in the human motor system. *J. Neurosci.* 19, 8043–8048.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., and Behrens, T. E. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.

Ringo, J. L. (1991). Neuronal interconnection as a function of brain size. *Brain Behav. Evol.* 38, 1–6.

Rizzolatti, G., and Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Exp. Brain Res.* 153, 146–157.

Rolheiser, T., Stamatakis, E. A., and Tyler, L. K. (2011). Dynamic processing in the human language system: synergy between the arcuate fascicle and extreme capsule. *J. Neurosci.* 31, 16949–16957.

Romanski, L. M., and Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nat. Neurosci.* 5, 15–16.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999a). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136.

Romanski, L. M., Bates, J. F., and Goldman-Rakic, P. S. (1999b).

Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 403, 141–157.

Rosazza, C., Cai, Q., Minati, L., Paulignan, Y., and Nazir, T. A. (2009). Early involvement of dorsal and ventral pathways in visual word recognition: an ERP study. *Brain Res.* 1272, 32–44.

Ross, E. D. (2010). Cerebral localization of functions and the neurology of language: fact versus fiction or is it something else? *Neuroscientist* 16, 222–243.

Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M. S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., and Weiller, C. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040.

Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Kupper, H., Kellmeyer, P., Kummerer, D., Kloppel, S., Glauche, V., Lange, R., Mader, W., Feess, D., Timmer, J., and Weiller, C. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *Neuroimage* 49, 3187–3197.

Schenker, N. M., Buxhoeveden, D. P., Blackmon, W. L., Amunts, K., Zilles, K., and Semendeferi, K. (2008). A comparative quantitative analysis of cytoarchitecture and minicolumnar organization in Broca's area in humans and great apes. *J. Comp. Neurol.* 510, 117–128.

Schenker, N. M., Desgouttes, A. M., and Semendeferi, K. (2005). Neural connectivity and cortical substrates of cognition in hominoids. *J. Hum. Evol.* 49, 547–569.

Schlaug, G., Marchina, S., and Norton, A. (2009). Evidence for plasticity in white-matter tracts of patients with chronic Broca's aphasia undergoing intense intonation-based speech therapy. *Ann. N.Y. Acad. Sci.* 1169, 385–394.

Schmahmann, J. D., and Pandya, D. (2006). *Fiber Pathways of the Brain.* Oxford: Oxford University Press.

Schneider, G. E. (1969). Two visual systems. *Science* 163, 895–902.

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., and Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum. Brain Mapp.* 32, 771–783.

Scott, S. K., and Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26, 100–107.

Scott, S. K., and Wise, R. J. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92, 13–45.

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123(Pt 12), 2400–2406.

Seltzer, B., and Pandya, D. N. (1984). Further observations on parieto-temporal connections in the rhesus monkey. *Exp. Brain Res.* 55, 301–312.

Shadmehr, R., and Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Exp. Brain Res.* 185, 359–381.

Sharp, D. J., Awad, M., Warren, J. E., Wise, R. J., Vigliocco, G., and Scott, S. K. (2010). The neural response to changing semantic and perceptual complexity during language processing. *Hum. Brain Mapp.* 31, 365–377.

Sherwood, C. C., Wahl, E., Erwin, J. M., Hof, P. R., and Hopkins, W. D. (2007). Histological asymmetries of primary motor cortex predict handedness in chimpanzees (Pan troglodytes). *J. Comp. Neurol.* 503, 525–537.

Stephan, K. M., Fink, G. R., Passingham, R. E., Silbersweig, D., Ceballos-Baumann, A. O., Frith, C. D., and Frackowiak, R. S. (1995). Functional anatomy of the mental representation of upper extremity movements in healthy subjects. *J. Neurophysiol.* 73, 373–386.

Szentagothai, J. (1983). The modular architectonic principle of neural centers. *Rev. Physiol. Biochem. Pharmacol.* 98, 11–61.

Szymanski, M. D., Bain, D. E., and Henry, K. R., (1995). "Auditory evoked potentials of a killer whale (Orcinus orca)," in *Sensory systems of aquatic mammals*, eds R. A. Kastelein, J. A. T., and P. E. Nachtigall (Woerden: De Spil), 1–10.

Tallal, P., and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends Neurosci.* 29, 382–390.

Tettamanti, M., and Weniger, D. (2006). Broca's area: a supramodal hierarchical processor? *Cortex* 42, 491–494.

Thiebaut de Schotten, M., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D. G., and Catani, M. (2011). A lateralized brain network for visuospatial attention. *Nat. Neurosci.* 14, 1245–1246.

Thiebaut de Schotten, M., Dell'Acqua, F., Valabregue, R., and Catani, M. (2012). Monkey to human comparative anatomy of the frontal lobe association tracts. *Cortex* 48, 82–96.

Thierry, G., Cardebat, D., and Demonet, J. F. (2003a). Electrophysiological comparison of grammatical processing and semantic processing of single spoken nouns. *Brain Res. Cogn. Brain Res.* 17, 535–547.

Thierry, G., Giraud, A. L., and Price, C. (2003b). Hemispheric dissociation in access to the human semantic system. *Neuron* 38, 499–506.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–293.

Tiwari, V. N., Jeong, J. W., Wilson, B. J., Behen, M. E., Chugani, H. T., and Sundaram, S. K. (2012). Relationship between aberrant brain connectivity and clinical features in Angelman Syndrome: a new method using tract based spatial statistics of DTI color-coded orientation maps. *Neuroimage* 59, 349–355.

Tootell, R. B., Dale, A. M., Sereno, M. I., and Malach, R. (1996). New images from human visual cortex. *Trends Neurosci.* 19, 481–489.

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychol. Forsch.* 31, 299–348.

Umarova, R. M., Saur, D., Kaller, C. P., Vry, M. S., Glauche, V., Mader, I., Hennig, J., and Weiller, C. (2011). Acute visual neglect and extinction: distinct functional state of the visuospatial attention system. *Brain* 134, 3310–3325.

Umarova, R. M., Saur, D., Schnell, S., Kaller, C. P., Vry, M. S., Glauche, V., Rijntjes, M., Hennig, J., Kiselev, V., and Weiller, C. (2010). Structural connectivity for visuospatial attention: significance of ventral pathways. *Cereb. Cortex* 20, 121–129.

Ungerleider, L. G., and Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165.

Vallar, G., Di Betta, A. M., and Silveri, M. C. (1997). The phonological short-term store-rehearsal system: patterns of impairment and neural correlates. *Neuropsychologia* 35, 795–812.

Vates, G. E., and Nottebohm, F. (1995). Feedback circuitry within a song-learning pathway. *Proc. Natl. Acad. Sci. U.S.A.* 92, 5139–5143.

Vidorreta, J. G., Garcia, R., Moritz-Gasser, S., and Duffau, H. (2011). Double dissociation between syntactic gender and picture naming processing: a brain stimulation mapping study. *Hum. Brain Mapp.* 32, 331–340.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., and Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30, 1414–1432.

Vry, M. S., Saur, D., Rijntjes, M., Umarova, R., Kellmeyer, P., Schnell, S., Glauche, V., Hamzei, F., and Weiller, C. (2012). Ventral and dorsal fiber systems for imagined and executed movement. *Exp. Brain Res.* 219, 203–216.

Vygotzky, L. S., (1934). *Thought and Language.* Cambridge, MA: MIT Press.

Vygotzky, L. S., (1978). *Mind in Society.* Cambridge, MA: Harvard University Press.

Wager, T. D., and Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 255–274.

Watkins, O. C., and Watkins, M. J. (1980). The modality effect and echoic persistence. *J. Exp. Psychol. Gen.* 109, 251–278.

Weiller, C., Bormann, T., Saur, D., Musso, M., and Rijntjes, M. (2011). How the ventral pathway got lost: and what its recovery might mean. *Brain Lang.* 118, 29–39.

Weiller, C., Isensee, C., Rijntjes, M., Huber, W., Muller, S., Bier, D., Dutschka, K., Woods, R. P., Noth, J., and Diener, H. C. (1995). Recovery from Wernicke's aphasia: a positron emission tomographic study. *Ann. Neurol.* 37, 723–732.

Weiller, C., Musso, M., Rijntjes, M. and Saur, D. (2009). Please don't underestimate the ventral pathway in language. *Trends Cogn. Sci.* 13, 369-70; 370–371.

Wernicke, C. (1874). *Der aphasische Symptomenkomplex. Eine psychologische Studie auf Anatomischer Basis.* Breslau: Cohn und Weigert.

Wernicke, C. (1906). "Der aphasische Symptomenkomplex," in *Deutsche Klinik am Eingang des zwanzigsten Jahrhunderts in akademischen Vorlesungen.* ed E. V. Leyden (Berlin: Urban, Schwarzenberg), 487–556.

Wiley, N. (2006). Inner Speech as a Language: a Saussurean Inquiry. *J. Theory Soc. Behav.* 36, 319–341.

Williams, E. L., and Casanova, M. F. (2010). Autism and dyslexia: a spectrum of cognitive styles as defined by minicolumnar morphometry. *Med. Hypotheses* 74, 59–62.

Wilson, B. J., Sundaram, S. K., Huq, A. H., Jeong, J. W., Halverson, S. R., Behen, M. E., Bui, D. Q., and Chugani, H. T. (2011). Abnormal language pathway in children with Angelman syndrome. *Pediatr. Neurol.* 44, 350–356.

Wilson, F. A., Scalaidhe, S. P., and Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* 260, 1955–1958.

Winsler, A., and Naglieri, J. (2003). Overt and covert verbal problem-solving strategies: developmental trends in use, awareness, and relations with task performance in children aged 5 to 17. *Child Dev.* 74, 659–678.

Winsler, A., Abar, B., Feder, M. A., Schunn, C. D., and Rubio, D. A. (2007). Private Speech and executive functioning among high-functioning children with autistic spectrum disorders. *J. Autism Dev. Disord.* 37, 1617–1635.

Winsler, A., Carlton, M. P., and Barry, M. J. (2000). Age-related changes in preschool children's systematic use of private speech in a natural setting. *J. Child Lang.* 27, 665–687.

Winsler, A., De Leon, J. R., Wallace, B. A., Carlton, M. P., and Willson-Quayle, A. (2003). Private speech in preschool children: developmental stability and change, across-task consistency, and relations with classroom behaviour. *J. Child Lang.* 30, 583–608.

Wise, R. J. (2003). Language systems in normal and aphasic human subjects: functional imaging studies and inferences from animal studies. *Br. Med. Bull.* 65, 95–119.

Wolpert, D. M., and Miall, R. C. (1996). Forward models for physiological motor control. *Neural Netw.* 9, 1265–1279.

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882.

Yeterian, E. H., Pandya, D. N., Tomaiuolo, F., and Petrides, M. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48, 58–81.

Zatorre, R. J., and Belin, P. (2001). *Spectral and Temporal Processing in Human Auditory Cortex.* Oxford, UK: Oxford University Press.

Zhang, J., Evans, A., Hermoye, L., Lee, S. K., Wakana, S., Zhang, W., Donohue, P., Miller, M. I., Huang, H., Wang, X., Van Zijl, P. C., and Mori, S. (2007). Evidence of slow maturation of the superior longitudinal fasciculus in early childhood by diffusion tensor imaging. *Neuroimage* 38, 239–247.

Zuberbuhler, K. (2002). A syntactic rule in monkey forest communication. *Anim. Behav.* 63, 293.

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; accepted: 11 June 2012; published online: 03 July 2012.

Citation: Rijntjes M, Weiller C, Bormann T and Musso M (2012) The dual loop model: its relation to language and other modalities. Front. Evol. Neurosci. ***4****:9. doi: 10.3389/fnevo.2012.00009*

frontiers in
EVOLUTIONARY NEUROSCIENCE

REVIEW ARTICLE
published: 01 February 2012
doi: 10.3389/fnevo.2012.00002

Gestures, vocalizations, and memory in language origins

Francisco Aboitiz**

Departamento de Psiquiatría, Facultad de Medicina y Centro Interdisciplinario de Neurociencia, Pontificia Universidad Católica de Chile, Santiago, Chile

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

Reviewed by:
Steven Chance, Oxford University, UK
Jonathan K. Burns, University of KwaZulu-Natal, South Africa

****Correspondence:***
Francisco Aboitiz, Departamento de Psiquiatría, Facultad de Medicina y Centro Interdisciplinario de Neurociencia, Pontificia Universidad Católica de Chile, Avenue Marcoleta #391, Santiago, Chile.
e-mail: faboitiz@puc.cl

This article discusses the possible homologies between the human language networks and comparable auditory projection systems in the macaque brain, in an attempt to reconcile two existing views on language evolution: one that emphasizes hand control and gestures, and the other that emphasizes auditory–vocal mechanisms. The capacity for language is based on relatively well defined neural substrates whose rudiments have been traced in the non-human primate brain. At its core, this circuit constitutes an auditory–vocal sensorimotor circuit with two main components, a "ventral pathway" connecting anterior auditory regions with anterior ventrolateral prefrontal areas, and a "dorsal pathway" connecting auditory areas with parietal areas and with posterior ventrolateral prefrontal areas via the arcuate fasciculus and the superior longitudinal fasciculus. In humans, the dorsal circuit is especially important for phonological processing and phonological working memory, capacities that are critical for language acquisition and for complex syntax processing. In the macaque, the homolog of the dorsal circuit overlaps with an inferior parietal–premotor network for hand and gesture selection that is under voluntary control, while vocalizations are largely fixed and involuntary. The recruitment of the dorsal component for vocalization behavior in the human lineage, together with a direct cortical control of the subcortical vocalizing system, are proposed to represent a fundamental innovation in human evolution, generating an inflection point that permitted the explosion of vocal language and human communication. In this context, vocal communication and gesturing have a common history in primate communication.

Keywords: arcuate fasciculus, broca's area, inferior parietal lobe, mirror neurons, phonological loop, superior longitudinal fasciculus, working memory

INTRODUCTION

In the last 15 years, there has been an increasing interest in understanding the evolutionary aspects of language and human communication. Several comparative analyses have been aimed at identifying a phylogenetic continuity between the brain networks involved in language processing in humans, and neural circuits present in the non-human primate. At least two lines of research have become particularly influential in this regard. One of them has focused on the search for auditory–premotor circuits in the macaque monkey, by assuming homology with the human's language network based on cytoarchitectonic and connectivity criteria (Aboitiz and García, 1997; Petrides and Pandya, 2009). These findings are broadly consistent with those obtained through a comparative approach, which studies vocal learning in non-human species, particularly in songbirds, as both emphasize the development of auditory–vocal circuits as a crucial step in the acquisition of human language (Bolhuis et al., 2010; Berwick et al., 2011).

Another research program emerged somewhat unexpectedly from the study of grasping visuomotor neurons in the parietal and premotor cortex of the monkey, where the so-called "mirror neurons" were found to be activated both when executing an action and when observing this action (Di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996). Based on these findings, Rizzolatti and Arbib (1998) developed the hypothesis that the grasping mirror neuron system represented a scaffold from which language circuits emerged in the human. Mirror neurons are found in area F5 of the ventral premotor cortex, which has been proposed by some authors to be the homolog of Broca's area in the human (Rizzolatti and Craighero, 2004).

These two approaches have largely been considered alternative possibilities, and there has been little cross-talk between authors supporting each view; in addition some misunderstanding of each other's work has increased the difficulty of reaching some agreement or common view. While the gestural and mirror neuron perspective makes strong emphasis on the background conditions for the emergence of human language, it does not provide specific insights into how speech arose to become the predominant communication mode in our species. In this article I will discuss some of the evidence supporting both views, in order to propose an integrated perspective in which the evolution of human communication has been based on multimodal signals including facial, hand, and body gestures, together with vocalizations.

AN EARLY HYPOTHESIS OF MONKEY HOMOLOGIES

ANCESTRY OF THE LANGUAGE CIRCUITS

Several years ago, we presented a hypothesis for the evolutionary origin of the language networks, based on the hodological evidence available for the monkey at that time (Aboitiz and García, 1997; see also Aboitiz, 1995). Homologs to human areas 44 and 45 (corresponding to Broca's region) in the ventrolateral prefrontal cortex (VLPFC) had been already identified in the chimpanzee by Brodmann, and there has been no further debate about their correspondence (Sherwood et al., 2003; Schenker et al., 2008; Keller et al., 2009). However, in the macaque the situation was more difficult. At that time, only area 45 could be identified in this species, inside the inferior arcuate sulcus, between subareas 6v and 8Ar (Preuss and Goldman-Rakic, 1991a). This region was viewed as a specialization of the premotor area 6v (area 6 ventralis), representing orofacial movements (Deacon, 1992; Preuss, 1995). On the other hand, area Tpt in the superior temporal lobe (which has been related to Wernicke's region by some authors) had been identified even in prosimians. This area was described as a multimodal zone receiving auditory and somatosensory projections from the temporal and parietal lobes, respectively (Galaburda and Sanides, 1980; Pandya and Yeterian, 1985; Preuss and Goldman-Rakic, 1991b).

Nonetheless, evidence for an arcuate fasciculus connecting Broca's and Wernicke's areas, was difficult to find in the monkey. Area Tpt was found to send projections to areas 8 and 46 of the prefrontal cortex, but not to area 45 or to the ventral arcuate sulcus (Petrides and Pandya, 1988). On the basis of these findings, area Tpt was proposed to play a role in head-turning movements aimed at localizing sound sources (Pandya and Yeterian, 1985). The only superior temporal projections to the ventral arcuate sulcus originated from the secondary auditory area ProA (Petrides and Pandya, 1988). However, another study at that time described some temporal projections from the superior temporal gyrus and the superior temporal sulcus (STS), to the inferior post-arcuate and the pre-arcuate region (Deacon, 1992). On the other hand, area 45 was found to receive projections from the anterior inferior parietal area 7b, which also sends afferents to area 46 in the dorsolateral prefrontal cortex (Petrides and Pandya, 1984; Preuss and Goldman-Rakic, 1991c; Seltzer and Pandya, 1994). Inside the intraparietal sulcus, area 7ip had been described as projecting to the dorsal and ventral aspects of the anterior arcuate sulcus (area 8), and to the posterior principal sulcus (area 46; Petrides and Pandya, 1984; Cavada and Goldman-Rakic, 1989; Preuss and Goldman-Rakic, 1991a,b,c). Pandya and Yeterian (1985) and Seltzer and Pandya (1978) described connections between the middle superior temporal lobe and somatosensory parietal regions via the middle longitudinal fasciculus (MLF), whereas Cavada and Goldman-Rakic (1989) reported projections from area 7ip to the posterior STS, and from areas 7b and 7a (the latter is posterior to 7b) to the superior bank of the STS. Finally, there is evidence for a projection from visual area TE in the inferior temporal lobe to the inferior arcuate sulcus, including area 45 (Bullier et al., 1996).

TRIPARTITE INPUT TO BROCA'S REGION

Based on these descriptions, we proposed – to our knowledge for the first time – a model for the organization of the language circuits and their possible homologies in the monkey, which emphasized a more complex network than did previous models. A significant component of this model was the inclusion of an inferior parietal projection to Broca's area and a connection between Wernicke's region and the inferior parietal lobe (Aboitiz and García, 1997). This schema implied three main inputs to Broca's area: (i) a direct route running through the arcuate fasciculus; (ii) an indirect route from the posterior superior temporal lobe to the inferior parietal lobe, and from there to Broca's region; and (iii) projections from the anterior temporal lobe to Broca's area (although we proposed that these were mainly visual; Aboitiz and García, 1997). In human evolution, the direct projections from Wernicke's area to Broca's area via the arcuate fasciculus would have gained greater importance than in the monkey, in which no clear evidence for an arcuate fasciculus existed. Furthermore, at the time several imaging studies had shown a strong inferior parietal involvement in verbal working memory, especially in phonological storage tasks (for example, Paulesu et al., 1993; Awh et al., 1996; Salmon et al., 1996; see also Smith and Jonides, 1998), which was consistent with an inferior parietal input to Broca's region. These findings were in line with our hypothesis that working memory, particularly phonological working memory, was important for language acquisition in children (Baddeley et al., 1988) and also in early humans. Nonetheless, following Fuster (1995), we also argued strongly that rather than there being specific memory-dedicated regions, short-term memory should be considered a property of the whole network involved in sensorimotor integration, which interacted intensively with other associated networks (Aboitiz and García, 1997; see also Aboitiz et al., 2006a,b, 2010). In summary, we proposed that an expansion of working memory capacity was critically associated with the differentiation of the above mentioned language circuits, in the context of learning and processing complex phonological sequences that were acquired by imitation of conspecifics (Aboitiz and García, 1997).

NEW EVIDENCE

AUDITORY PROJECTIONS IN NON-HUMAN PRIMATES

After our original publication, there has been a wealth of new evidence on the existence of temporal–parietal–prefrontal connections, both in the monkey and in the human. In the macaque, auditory projections separate into a dorsal and a ventral stream, running to the parietal lobe and to the anterior temporal lobe, respectively. This arrangement has been viewed as being analogous to the bipartite arrangement of the visual system, in which the dorsal pathway is involved in spatiotemporal signal processing and is related to eye movement control, whereas the ventral pathway relates to stimulus recognition and emotional processing (Kaas and Hackett, 1999).

In the macaque, the dorsal auditory pathway originates in posterior auditory areas in the superior temporal lobe and is directed mainly to dorsal prefrontal areas (areas 8 and 46, related to eye movement control). It is noteworthy that this pathway does not fit in an obvious manner into the language network, as it terminates preferentially in dorsal rather than ventral prefrontal regions (in the human, a dorsal prefrontal projection of the dorsal pathway has been also described; see Frey et al., 2008). On the other hand, the ventral stream originates in different areas of the anterior and middle temporal gyrus, and conveys visual and auditory inputs directed mainly to areas 12 and 45 of the VLPFC (Kaas and Hackett, 1999; Romanski et al., 1999a,b; Belin and Zatorre, 2000; Rauschecker and Tian, 2000; Romanski, 2007). Consistent with this evidence, other reports described an auditory domain in the macaque inferior frontal areas 12 and 45, in which vocalization-specific neurons were interspersed with facial-sensitive neurons, allowing for the integration of vocal auditory stimuli with the corresponding facial gestures (Romanski and Goldman-Rakic, 2002; Romanski et al., 2005; Romanski, 2007). Interestingly, this region was found to receive afferents from the anterior lateral belt auditory area (Rauschecker and Tian, 2000; Tian et al., 2001), which is preferentially activated by calls from conspecifics (Petkov et al., 2008).

INFERIOR PARIETAL PROJECTIONS TO THE VLPFC OF THE MACAQUE

Petrides and Pandya (1999, 2002) subdivided the monkey area 45 into areas 45A and 45B, and identified a dysgranular area 44 in the depth of the inferior arcuate sulcus (Petrides et al., 2005). Stimulation of neurons in area 44 triggered orofacial movements and sometimes hand movements, but not ocular movements; oculomotor responses occurred only when area 8Av was stimulated, far from the 44–8Av border (Petrides et al., 2005). Furthermore, stimulation sites in the most ventral aspect of area 8Av and in the 45–8Av border did not elicit any motor response.

Before discussing in more detail the different reports and interpretations on connectivity of the inferior parietal lobe, it must be noted that the cytoarchitectonic parcellation of this region has not been consistent across studies. Whereas earlier studies adopted Brodmann's early description of area 7, subdividing it into areas 7b and 7a, and area 7ip inside the intraparietal sulcus (Petrides and Pandya, 1984; Preuss and Goldman-Rakic, 1991a,b,c), more recent studies have used another parcellation scheme, defining area PF anteriorly (Brodmann's area 40, anterior supramarginal gyrus in the human), area PFG in the middle (area 39, posterior supramarginal gyrus in the human), and area PG posteriorly (area 39, angular gyrus in the human), with area anterior intraparietalis (AIP) inside the intraparietal sulcus (see Matelli et al., 1986; Frey et al., 2008; Petrides and Pandya, 2009; Gerbella et al., 2011).

Petrides and Pandya (1999, 2002) described area 45 as being connected with the posterior inferior parietal lobe, while area 44 was viewed as receiving projections from the intraparietal and anterior inferior parietal lobe. Subsequently, Petrides and Pandya (2009) visualized a pattern of multiple afferents into areas 45A, 45B, and 44 from the inferior parietal and temporal regions. These projections consist of two main pathways: (i) axons running along the SLF into both areas 45 and 44, and which originate in the inferior parietal lobe (areas PFG and PG). Area PFG made a particularly strong projection into area 44. In addition, they described some axons from the ventral most inferior parietal lobe and the caudal STS, which formed an arcuate fasciculus, although this projection is not as prominent as it is in humans. Furthermore, there was a systematic relation between inferior parietal regions and the prefrontal regions to which they connected, with more rostral parts (area PF) connecting with the ventral premotor cortex (area 6 ventralis, controlling facial musculature), while intermediate regions of the inferior parietal lobe (area PFG) connected to area 44 and to a lesser extent to area 45. The second pathway (ii) consisted of multimodal axons running via the extreme capsule and uncinate fasciculus, originating in diverse auditory and visual cortical areas of the anterior and middle temporal lobe, and ending mainly in areas 45 and 47/12, but also to some extent in area 44. These authors argue that, in both the monkey and in the human, the ventral projection to the VLPFC has a role in the mechanisms of memory retrieval, while the dorsal route (arcuate and superior longitudinal fasciculi) is involved in the control of vocal articulation only in humans (see also Saur et al., 2008).

TRACTOGRAPHIC STUDIES IN THE HUMAN BRAIN

Likewise, the advent of tractographic techniques in the living human yielded results consistent with the tripartite projection from the auditory regions into Broca's area that we originally described, with some modifications (Catani and ffytche, 2005; Parker et al., 2005; Friederici et al., 2006; Anwander et al., 2007; Frey et al., 2008; Glasser and Rilling, 2008; Friederici, 2009). Glasser and Rilling (2008) reported a two-component arcuate fasciculus in the left hemisphere, one connecting the superior temporal gyrus with areas 6 and 44, which according to them subserves phonological information; and the other connecting the middle temporal gyrus with areas 9, 44, and 45, and proposed to be involved in lexical-semantic aspects. In the right hemisphere they visualized a less prominent fasciculus, connecting the middle temporal gyrus with areas 6 and 44, which was proposed to convey prosodic information. They also reported a very small tract connecting the superior temporal lobe with areas 6 and 44 in the right hemisphere. Likewise, Parker et al. (2005) reported a strong asymmetry in the arcuate fasciculus, favoring the left hemisphere; a similar asymmetry was present in infants 1–4 months of age (Dubois et al., 2009). In a subsequent article, Rilling et al. (2008) visualized a progressive development of the arcuate fasciculus from human to macaque to the chimpanzee, while the ventral pathway, via the anterior temporal lobe, has remained more conservative during evolution (Rilling et al., 2008).

Frey et al. (2008) also described an arcuate fasciculus running from the posterior superior temporal gyrus to area 44 (in some cases to area 45; **Figure 1**). Note that the human arcuate fasciculus also projects to dorsal prefrontal areas 8 and 6, as in the monkey. However, the main focus of this report was on the inferior parietal and anterior temporal lobe projections to Broca's area. They found (i) a projection from the inferior parietal lobe (supramarginal gyrus) into area 44 via the SLF (in 10 of 12 subjects). The ventral posterior intraparietal region is claimed to receive auditory afferents from the superior temporal lobe via the middle and inferior longitudinal fasciculi, which might close a circuit from the posterior auditory cortex to area 44. In addition, they reported (ii) a ventral pathway connecting the anterior temporal areas with areas 47 and 45 via the extreme capsule and the uncinate fasciculus, as occurs in the monkey. This pathway has been described in other reports, and participates in the recognition of auditory stimuli including speech, identifying the speaker, mapping sound with meaning, verbal retrieval, echoic memory, and in simple grammatical processing (Buchsbaum et al., 2005a,b; Hickok and Poeppel, 2007; Saur et al., 2008). According to Hickok and Poeppel (2007) the ventral pathway, unlike the dorsal stream, is represented bilaterally, being functional in both the left and right hemispheres. Recent evidence has shown that, instead of being separate pathways, the dorsal and the ventral streams operate synergistically during language processing (Lopez-Barroso et al., 2011; Rolheiser et al., 2011).

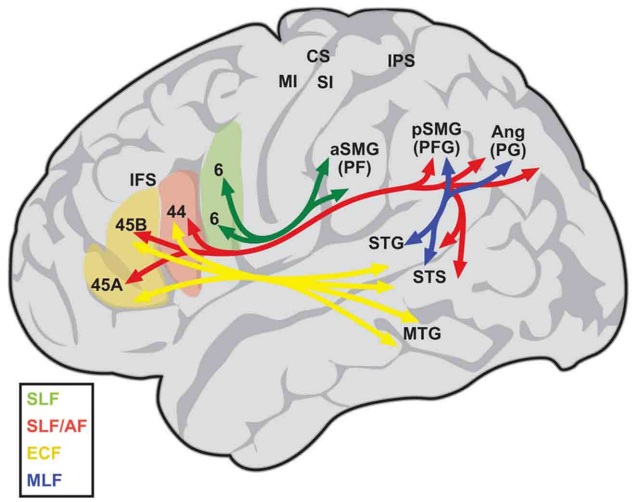


FIGURE 1 | Diagram depicting the language-related circuit in humans, as proposed by Frey et al. (2008). The superior longitudinal fasciculus (SLF) connects inferior parietal areas PF with the ventral premotor cortex (area 6; green), while areas PFG and PG are connected with areas 44 and 45 (red). The arcuate fasciculus (AF) connects posterior superior temporal regions with areas 44 and 45 as well (red), but is difficult to separate from the inferior branch of the SLF. The middle longitudinal fasciculus (MLF, blue) connects the posterior superior temporal gyrus and sulcus (STG, STS) with inferoparietal regions PFG and PG. Finally, a ventral route running via the extreme capsule (ECF, yellow) connects the middle and anterior temporal lobe with areas 44 and 45. A similar circuit has been described for the monkey (Petrides and Pandya, 2009). Ang, angular gyrus; aSMG, anterior supramarginal gyrus; CS, central sulcus; IPS, intraparietal sulcus; MI, primary motor area; MTG, middle temporal gyrus; pSMG, posterior supramarginal gyrus; SI, primary somatosensory area. Based on Kelly et al. (2010), with permission.

Despite the attractiveness of these studies, it is not entirely clear to what extent the tractographic evidence reveals a monosynaptic arcuate fasciculus, first because this technique lacks the resolution required to strongly confirm this possibility, and second, because this tract is difficult to separate from the adjacent SLF, which carries fibers from the parietal lobe. It must be also mentioned that Bernal and Altman (2010) were unable to find a strong projection from the SLF or the arcuate fasciculus into Broca's region, finding instead a strong termination of these tracts in the ventral premotor and motor cortices. However, in this study the parietal site of origin of the SLF was insufficiently characterized, which undermines somewhat the authors' main conclusions.

THE INFERIOR PARIETAL LOBE AND VERBAL WORKING MEMORY

There is now an important discussion about the role of different inferior parietal areas in verbal working memory, as the concept of a memory-dedicated, anatomically isolated component has been challenged by many studies (Hickok and Poeppel, 2007; Buchsbaum and D'Esposito, 2008; Hickok, 2009; see also Aboitiz et al., 2006a, 2010). More importantly, the only regions that have consistently shown sustained activation during verbal working memory tasks are the STS and the mid-superior temporal gyrus, especially an area located in the posterior planum temporale (area Spt; its relation to area Tpt is not yet clear; Buchsbaum et al., 2005a,b; Hickok and Poeppel, 2007; Hickok, 2009). In particular, area Spt is proposed to serve as an interface between sensory and motor representations during the maintenance of phonological items on line (Buchsbaum and D'Esposito, 2008; Buchsbaum et al., 2011). In this interpretation, the "phonological loop" is considered to include a sensory phonological processing system partly represented (bilaterally) by the STS, a sensory–motor integration system in the left Spt area, and a left frontal articulatory system (Hickok, 2009). These authors interpret the role of the inferior parietal lobe as serving some higher-order functions that support verbal working memory. One possibility is that these regions participate in motor planning mechanisms that help stabilize perceptual memory traces (see below).

THE PHONOLOGICAL LOOP: A KEY INNOVATION

PHONOLOGICAL CIRCUITS AND WORKING MEMORY

In subsequent reports, we emphasized the role of phonological working memory, supported by the development of a phonological sensory motor circuit (the phonological loop) in early humans, as a crucial element in early language and human evolution (Aboitiz et al., 2005, 2010). The ability to rehearse and keep newly learned phonological sequences in short-term memory became an inflection point that changed human sociality forever, being a fundamental factor in the evolution of complex language and culture. This "inner speech" capacity also allowed the elaboration of new and more complex messages by manipulating the phonemes being learned. Furthermore, we argued that this circuit was largely, although not exclusively, based on the strengthening of the dorsal pathway connecting Wernicke's and Broca's areas (including the direct projection via the arcuate fasciculus and the indirect one via the inferior parietal lobe), while the ventral pathway running via the extreme or external capsule had been more conservative in evolution (although not static; see below), being the dominant pathway for vocalization processing in the monkey (Aboitiz et al., 2006a, 2010).

SYNTAX

We also claimed that the origin of a complex syntax based on long-distance dependencies between linguistic elements required a robust phonological memory system in order to maintain the different items active while other elements were being processed online (Aboitiz et al., 2006a, 2010). Imaging studies indicate a participation of Broca's area in working memory processes associated with syntactical processing (Fiebach et al., 2002, 2005), and Friederici (2004) has argued that syntactic working memory involves the superior anterior portion of area 44, while syntactic processing relates to the inferior portion of left area 44. Furthermore, the dorsal pathway for language is involved in the processing of grammatical structures organized in a hierarchical manner, whereas a "middle pathway," similar to the ventral pathway but ending in area 45A, participates in the analysis of simple grammatical structures (Friederici et al., 2006; Anwander et al., 2007; see also Wilson et al., 2011).

In the adult, syntactical processing is probably automatic to a large extent – especially simple grammatical forms – and may depend on cortico-striatal circuits involved in procedural memory

(Ullman, 2004). Along the same line, patients with lesions in the left temporo-parietal cortex that have specific short-term memory deficits for numbers and words, do not display any major impairments in their spontaneous speech, supporting the participation of subcortical components in automatic language processing (Shallice and Warrington, 1970; Saffran and Marin, 1975). However, the initial acquisition of rules, the processing of complex syntactic forms (Friederici, 2005), and the online maintenance of semantic information during linguistic processing, require short-term memory mechanisms that bridge these procedural components with episodic memory networks (Reuland, 2010).

HANDEDNESS, GESTURES, AND MIRROR NEURONS

BRAIN ASYMMETRIES FOR LANGUAGE AND HAND CONTROL

A critical issue in the context of language evolution is the conspicuous left hemispheric specialization for linguistic functions in most people, which is consistent with the evidence of gross-anatomical asymmetries in language-related regions (Ide et al., 1999; Josse and Tzourio-Mazoyer, 2004). Interestingly, asymmetry for language is correlated with some lateralized capacities like handedness, but not with other asymmetric capacities such as spatial attention (Badzakova-Trajkov et al., 2010). Apes tend to be right-handed, and there is evidence suggesting that Neanderthals were predominantly right-handed as well (Lonsdorf and Hopkins, 2005; Steele and Uomini, 2009). Notably, in chimpanzees, there is a correlation between throwing capacity, communicative ability and, the white-to-gray matter ratio in the homolog of Broca's area (Hopkins et al., 2012).

Thus, handedness, throwing ability, and rhythmic hammering have been related to language origins, which is in line with the mirror neuron hypothesis (see below; Calvin, 1983; Corballis, 2003). Along this line, several authors have made emphasis on gestural and manual communication as a first step in the acquisition of language (Hewes, 1973; Corballis, 1992; Armstrong et al., 1995; Kendon, 2004). More specifically, Corballis (1992, 2002) originally proposed that generativity, a key syntactic operation, was initially present in a system of manual gestures, but switched to a predominantly vocal system in modern humans. Corballis included evidences from different fields of comparative cognition and the mirror neuron literature in his hypotheses, which strictly imply a stage of predominantly manual communication before vocal language took over (Corballis, 2002, 2003, 2010).

A more general, but not alternative, interpretation of hemispheric dominance for language is that complex sequential motor patterns may be more efficiently programmed in one hemisphere than in two. This fits with comparative evidence of lateralization for song production in songbirds (Bolhuis et al., 2010). Other authors have proposed that differences in interhemispheric communication via the corpus callosum may have played a role in the origin or maintenance of human brain lateralization (Ringo et al., 1994; Aboitiz et al., 2003; Häberling et al., 2011).

THE DISCOVERY OF MIRROR NEURONS

While studying the neurophysiology of visuomotor neurons involved in hand-grasping control in the monkey, Di Pellegrino et al. (1992) observed a group of motor neurons, termed "mirror neurons" that also became active when the animal observed meaningful hand movements made by the experimenters or by another animal (see also Rizzolatti and Luppino, 2001; Rizzolatti and Craighero, 2004). Most mirror neurons were initially observed in the premotor area F5 (Brodmann's area 6v), located in the precentral gyrus and adjacent to the inferior arcuate sulcus (Belmalih et al., 2009). Area F5 has been subdivided into areas F5p, F5c, and F5a. Area F5a, which is adjacent to area 44 in the inferior arcuate sulcus (see below) has been proposed to be an integration site for parietal sensory–motor signals with signals from prefrontal and premotor areas (Gerbella et al., 2011). Of note, mirroring properties were also observed in face-selective neurons of the lateral aspect of F5, possibly allowing the animal to recognize gestures produced by conspecifics. Many of these responded to feeding behaviors, but some also fired when the animal observed a communicative gesture like a lip smacking (Ferrari et al., 2003; Rizzolatti and Craighero, 2004). Furthermore, some mirror neurons were found to fire not only in response to an observed action, but to action-related sounds, even in the absence of the visual presentation of the action (Keysers et al., 2003).

Mirror neurons have also been described in the rostral inferior parietal area, firing both to the observation of actions and to the execution of these or similar actions (Fogassi et al., 1998; Gallese et al., 2002). In the STS (which is connected with the inferior parietal region), there are sensory neurons selective for body actions rather than to grasping, although some of them also fire with the observation of goal-directed hand movements (Perrett et al., 1990). As seen by fMRI in the monkey, observation of grasping actions produces activations in inferior frontal areas F5, 45B, 45A, and 46; and on parietal areas PFG and AIP, plus the STS (Nelissen et al., 2005, 2011). These authors suggest that there are two pathways involved in the observation of actions, one running from the upper STS, relaying in area PFG, and projecting to the premotor area F5c which processes the agent's intentions (a context-dependent representation of the action); and the other, that originates in the lower STS, projects to area F5a/p via AIP and is more focused on the object (**Figure 2**; see also Luppino et al., 1999). They also describe connections of the STS and the lateral intraparietal area with area 45B. In the monkey, area PF was found to project strongly to F5a, F5c, and F5p; area PFG directed its axons to area F5a and F5p, while area PG was mainly directed to area F5p (Matelli et al., 1986; Gerbella et al., 2011). In other words, area F5a receives a robust input from areas PFG and AIP, a weaker input from PF, and practically no input from area PG (Gerbella et al., 2011). In addition, AIP projects to premotor area F5 (representing mostly the hand and mouth), while the ventral intraparietal sulcus is connected with the more dorsal premotor area F4 (representing the arm, neck, and face; Luppino et al., 1999). Note that this pattern is different from that described in the macaque by Petrides and Pandya (2009), who emphasize inferior parietal projections into the more anterior VLPFC.

A MIRROR SYSTEM IN HUMANS?

For obvious reasons, mirror neurons have been difficult to report in humans (Rizzolatti and Craighero, 2004). Nonetheless, there is a wealth of stimulation, electroencephalographic, and imaging data that is consistent with the notion that a mirror neuron system, i.e., a network involved in action recognition, imitation, and empathic

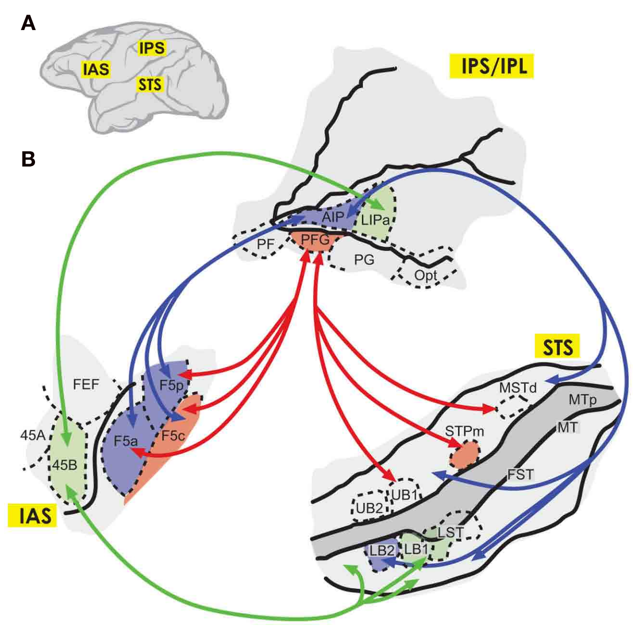


FIGURE 2 | (A) Diagram depicting the location of the inferior arcuate sulcus (IAS), the intraparietal sulcus and inferior parietal lobe (IPS and IPS/IPL), and the superior temporal sulcus (STS) in the macaque brain. **(B)** Pathways involved in action understanding, according to Nelissen et al. (2011). In red, an intention-processing pathway connecting the upper (STS) with area PFG and mainly frontal area F5c; and in blue, an object-related pathway connecting the lower STS with area AIP and areas F5p and F5a. There are also connections to area 45B from area lateral intraparietal area a (LIPa) and the anterior STS (green). No projections are shown here from areas PFG and PF, but related studies have described projections from area PFG into F5a and F5p; and from PG to F5p (Gerbella et al., 2011). Modified from Nelissen et al. (2011), with permission.

behavior, is present in the human (Rizzolatti and Craighero, 2004; Iacoboni and D'Apretto, 2006). However, there is discussion about whether this activity reflects or not the activity of mirror neurons as described in the monkey, and whether the human mirror neuron system does actually participate in language processing (Molenberghs et al., 2009; de Zubicaray et al., 2010). Below I will address some of the main findings of this research program, which nonetheless bears relevance to the issue of language and gesture interaction.

Unlike the monkey, humans show mirror-system activity with the observation of meaningless, not object-directed movements, and with pantomimes, which may be attributed to communication skills (Fadiga et al., 1995; Buccino et al., 2001; Maeda et al., 2002; Grèzes et al., 2003). In humans, the localization of mirror system activity encompasses a wide bilateral cortical network, including parietotemporal visual regions, the rostral inferior parietal lobe, and the inferior precentral and frontal gyri (Iacoboni and D'Apretto, 2006). More recent proposals also emphasize the participation of a ventral pathway running via the anterior temporal lobe, as an additional component involved in planning, decision making (Arbib, 2010), and in the prediction of the intentions and the goals of actions (Kilner, 2011).

In humans, the mirror system has been interpreted as participating in action understanding, which is critical for inferring another's intentions in a social context (Rizzolatti and Craighero, 2004). There are many studies that have reported an activation in Broca's region during real and imagined hand movements (Binkofski et al., 1999; Iacoboni et al., 1999; Gerardin et al., 2000). Furthermore, activation of area 44 with object-related mouth movements and imitation of vocal gestures has been reported in several studies (Di Pellegrino et al., 1992; Buccino et al., 2001). In addition, the pars triangularis, corresponding to area 45, displays mirror activity with the observation of behavioral goals rather than with the action itself (Johnson Frey et al., 2003). Finally, the mirror system has been shown to be involved in imitation tasks (Iacoboni and D'Apretto, 2006). During a finger imitation task in humans, Iacoboni et al. (1999) found a specific activation of the left pars opercularis (area 44), while in a task requiring the learning of a motor sequence, the activated areas included the pars opercularis, ventral premotor area, and the STS (Buccino et al., 2004; Vogt et al., 2007).

MIRROR NEURONS AND LANGUAGE CIRCUITS

MIRROR NEURONS AS A REQUIREMENT FOR LANGUAGE

On the basis of these and other findings, Rizzolatti and Arbib (1998) and Arbib (2005) proposed the bold hypothesis that the neural circuits involved in language processing evolved as an elaboration of the mirror neuron circuitry present in monkeys, which provided a scaffolding for the elaboration of a more complex, phonological network involved in communication and eventually, in speech. Furthermore, and as we originally claimed (Aboitiz and García, 1997), imitation is a key element in learned communication, and mirror neurons provide an adequate neural substrate for its implementation (even if monkeys are not good imitators). More specifically, Arbib (2005) proposed a sequence of events starting with an imitation system for grasping, which developed into a complex gestural communication system in which pantomime came to be used as a conventionalized reference system. Afterward, a "protosign" stage that used hand symbols would have occurred that eventually incorporated vocal sounds, or "protospeech." Likewise, words resembling or suggesting ingestive behavior were proposed to be particularly important for the origin of a primitive semantics (Ferrari et al., 2003; Rizzolatti and Craighero, 2004). More recently, Arbib (2010) proposed that the ventral pathway for actions may have been particularly relevant for the acquisition of a primitive semantics, as this may have evolved to support words-as-phonological-actions, with semantics provided by the linkage to neural systems supporting perceptual and motor schemas. This view is consistent with the current understanding of the ventral pathway as being involved in the transformation of sound into meaning (Buchsbaum et al., 2005a).

WHERE IS THE MONKEY HOMOLOG OF BROCA'S AREA?

Initially, proposers of the mirror neuron hypothesis identified area F5 as the most likely homolog of the human Broca's area (Rizzolatti and Craighero, 2004). More recently, Gerbella et al. (2007, 2010) confirmed Petrides and Pandya's (2002, 2009) descriptions of the monkey VLPFC, but emphasized connectivity of area 45A and 45B with oculomotor regions. They found only weak connections between area 45A and the inferior parietal areas PFG and PG, and between area 45B and the lateral intraparietal area. Thus, area 45A might be associated with eye movement control during communication, while area 45B would instead belong to the monkey prearcuate region, involved with other oculomotor processes. Instead, Petrides and Pandya (2009) assert that the presumed part of area 45 that has been linked to oculomotor function, cytoarchitectonically corresponds to the caudal oculomotor area 8. Gerbella et al. (2007, 2010) also confirmed the existence of area 44 in the monkey, and considered it as an anterior subdivision of area F5 (or area 6v), adjacent to area F5a (Belmalih et al., 2009; **Figure 3**). According to these authors, in their original description Petrides and Pandya (2002) described area 44 with a more posterior extension, overlapping with the anterior premotor area (F5), but they subsequently restricted the limits of this area to the actual fundus of the inferior arcuate sulcus (Petrides et al., 2005). This points to the concept of area 44 as a specialization of the ventral premotor area.

DISCUSSION

Much of the disagreement between scholars invoking auditory–vocal vs. hand-based ancestral circuits for language can be separated into two main issues: one concerns the correspondence in primates of the language-related circuits in humans, which focuses on identifying the cortical area ancestral to Broca's region in the monkey; and the other refers to the possibility that a specific hand-gestural communication system preceded the advent of speech, and on the likelihood that a hand-based mirror neuron system represents a critical scaffolding for the subsequent evolution of language.

HOMOLOGY ISSUES

The unequivocal identification of areas 44 and 45 in the macaque, in the chimpanzee, and in the human, with practically identical topographies and cytoarchitectonic features suggests that those areas are most likely homologous to each other, deriving from the same germinal field in the embryonic telencephalon. On the other hand, classical grasping or mouth mirror neurons have been located in the premotor area F5 (area 6 ventralis), near the border with area 44, in which there is yet no evidence of mirror neurons. However, it would be interesting to revisit the location of orofacial mirror neurons according to this cytoarchitectonic scheme (Ferrari et al., 2003), as in the monkey, stimulation of area 44 has been shown to elicit oral movements, either during communication or in feeding (Petrides et al., 2005).

Area 45 (and the adjacent area 12) fits the prefrontal auditory domain, receiving multimodal projections from the mid- and anterior-temporal lobe (Romanski, 2007; Petrides and Pandya, 2009; Gerbella et al., 2010). While Petrides and Pandya (2009) claim that this area participates in memory (semantic) retrieval processes, Belmalih et al. (2009) argue for a role in communication-directed eye movements, especially area 45B (see also Leichnetz, 2001). Although these discrepancies need to be resolved, it may safely be stated that the multimodal arrangement of area 45 corresponds to an auditory–motor interface that may be the evolutionary precursor of a speech-specialized region.

IS THE ARCUATE FASCICULUS THERE?

In the human left hemisphere, Glasser and Rilling (2008) described a tract connecting the superior temporal gyrus with areas 6 and 44 (involved in phonological processing), and a more robust one connecting more inferior temporal areas with areas 44, 45, and 9 (involved in semantic and lexical aspects). Frey et al. (2008) in the human, and Petrides and Pandya (2009) in the monkey, identified an arcuate fasciculus originating in the STS or in the inferior most parietal lobe, directed to the VLPFC and dorsal prefrontal areas. In the monkey, Yeterian et al. (2012) have recently made a claim for the existence of a direct projection between area Tpt in the superior temporal lobe and areas 44, 45, and dorsal prefrontal areas via the arcuate fasciculus. Furthermore, recent studies of verbal working memory point to the superior temporal gyrus (area Spt) as a key element involved in phonological sensorimotor integration (Buchsbaum and D'Esposito, 2008), which may perhaps contribute fibers to the arcuate fasciculus. Thus, the arcuate fasciculus is possibly an element involved in

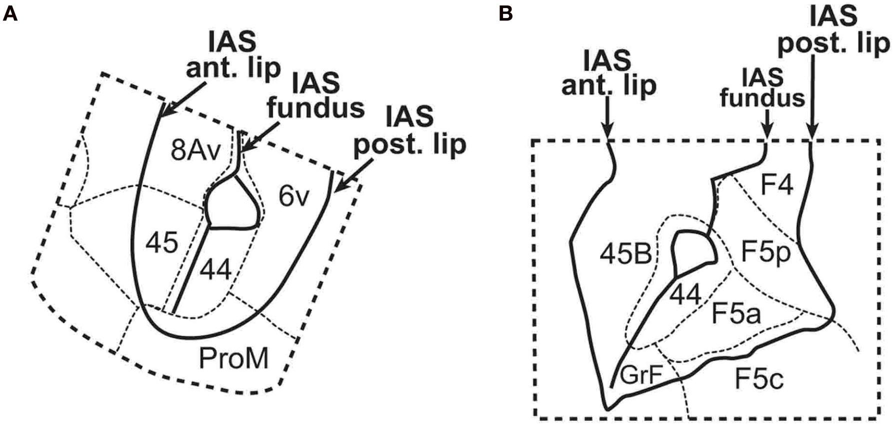


FIGURE 3 | Parcellation of the inferior arcuate sulcus (IAS) of the monkey according to (A) Petrides et al. (2005), and (B) Belmalih et al. (2009). In both cases, area 44 is shown in the depth of the sulcus, bordered anteriorly by area 45, and posteriorly by the premotor area (6v or F5a depending on the nomenclature). IAS, inferior arcuate sulcus; IPS/IPL, intraparietal sulcus and inferior parietal lobe, respectively; STS, superior temporal sulcus. From Belmalih et al. (2009), with permission.

auditory–vocal coordination and articulatory control, and might be involved in working memory processes by maintaining the functional connectivity between sensory and motor regions while holding items online. Nonetheless, tractographic studies to date are still insufficient to determine to what extent this is a monosynaptic pathway, different from the SLF, connecting posterior temporal and VLPFC regions. In the monkey, chemical tracing studies suggest that if it is present, it is rather small (Petrides and Pandya, 2009).

THE INFERIOR PARIETAL CONNECTION

According to several studies in monkeys, areas 45 and 44 receive strong or moderate afferences from the inferior parietal lobe (Mesulam et al., 1977; Petrides and Pandya, 1984, 1999, 2002, 2009; Cavada and Goldman-Rakic, 1989; Preuss and Goldman-Rakic, 1991c; Leichnetz, 2001). However, in other studies only minor inferior parietal and intraparietal projections were described into areas 45A and 45B, respectively (Belmalih et al., 2009; Gerbella et al., 2010). More posteriorly, the premotor area 6v (or F5, where mirror neurons have been detected) receives strong projections from inferior parietal and intraparietal areas (Petrides and Pandya, 2009; Gerbella et al., 2011; Gharbawie et al., 2011). In humans, an inferoparietal projection to areas 44 and 45 has been described in several tractography studies (Catani and ffytche, 2005; Parker et al., 2005; Friederici et al., 2006; Anwander et al., 2007; Frey et al., 2008; Friederici, 2009; but see Bernal and Altman, 2010). Additional and more extensive connectivity studies are needed to determine the exact pattern of inferior parietal–prefrontal projections in the monkey and in the human.

An additional pathway involved in this circuit consists of a projection via the MLF to the inferior parietal lobe and intraparietal sulcus, which originates in the superior temporal lobe and STS. For some authors, this projection carries auditory information (Keysers et al., 2003; Frey et al., 2008; Petrides and Pandya, 2009), whereas others consider it as conveying body and arm positional information (Luppino et al., 1999; Nelissen et al., 2011). Both interpretations are not necessarily exclusive, as this projection likely transmits a multimodal input to the inferior parietal lobe.

A participation of inferior parietal regions in language circuits has been acknowledged by several researchers, although the precise role of these areas has yet to be resolved (Buchsbaum and D'Esposito, 2008; see above). Some authors have proposed a relation to phonological processing (Moser et al., 2009; Hartwigsen et al., 2010; Turkeltaub and Coslett, 2010), while others propose a role restricted to the sensorimotor control of writing (Brownsett and Wise, 2010). According to some authors, the inferior parietal lobe participates at an interface between speech audition and the articulatory code (Hickok, 2009; Moser et al., 2009). As mentioned above, one possibility is that these circuits maintain the stability of phonological sensorimotor circuits by codifying motor plans that contribute to maintain a behavioral goal during a working memory task. Along this line, the inferior parietal cortex of primates, and its projections into the frontal cortex, codify a diversity of orienting and object-directed behaviors, and have been proposed to participate in the selection of appropriate actions among competing circuits (Gharbawie et al., 2011; Kaas et al., 2011). At some point in hominid evolution, these projections may have come to receive an increasingly stronger auditory input (especially phonological information from the STS), via the MLF, recruiting regions that were involved in face and especially mouth control to process vocalization information, and to perform action selection based on auditory input. It is very likely that this transition was concomitant with the elaboration of the direct cortical control over the hypoglossal motoneurons involved in vocalization (Jürgens and Alipour, 2002), thus closing a sensorimotor pathway from the acoustic system to the phonatory effectors. The recent finding of neurons controlling voluntary vocalizations in the ventral premotor cortex of the macaque is of great interest in this context, as it indicates that this rudimentary circuit was present at very early stages, possibly overlapping with other voluntary control systems (Coudé et al., 2011).

Finally in this section, the ventral pathway from the anterior temporal lobe to the anterior VLPFC (areas 45 and 47/12) has apparently suffered less structural changes in the lineage leading to humans (Rilling et al., 2008), which is consistent with our early hypothesis that in monkeys the ventral pathway is the dominant circuit involved in auditory–vocal integration (Aboitiz et al., 2006a). Nonetheless, in the human this pathway has been proposed to contribute to the processing of semantic, echoic information, and simple grammatical forms (Buchsbaum et al., 2005a; Anwander et al., 2007; Frey et al., 2008), indicating that it has suffered important modifications as well. It is also important to note that, despite being organized in several parallel streams, like the visual pathways, language processing operates in an integrated dynamics, in which all these streams converge on the common bottleneck of Broca's area, and very likely there is cross-pathway communication along the different functional routes (Rolheiser et al., 2011).

A MULTIMODAL COMMUNICATION SYSTEM

As discussed above, there is a confluence of facial gesture and vocalization information in the VLPFC of the monkey, mostly carried by the ventral visual and auditory pathways. Associations between the vocalization-sensitive region described in areas 45 and 47/12 (Romanski, 2007), the facial gesture-coding area 44 (Petrides et al., 2005), and the hand and body representations in the premotor area F5 are supported by neuroanatomy and make it plausible to visualize an integrated processing of hand and face gestures and vocalization patterns. Chimpanzees are able to match vocalizations with gesturing faces (Izumi and Kojima, 2004), and the chimpanzee homolog of Broca's area becomes active during both gestural and vocal communicative actions (Taglialatela et al., 2008); activation is maximal when gestures are accompanied with vocalizations to call the other's attention (Taglialatela et al., 2011). In humans, areas 44, 45, and 47 become activated during the integration of speech with gestures (Willems et al., 2007; Gentilucci and Dalla Volta, 2008), and there is evidence for activation of hand motor systems during speech (Gentilucci et al., 2001; Meister et al., 2003). Thus, communication is multimodal both in humans and monkeys, and makes use of overlapping circuits in both species (Aboitiz and García, 2009). This evidence supports the concept that early steps of language evolution also consisted of multimodal signals, instead of being predominantly hand-based or vocalization-based.

WERE GESTURES OR GRASPING REQUIRED FOR THE ADVENT OF SPEECH?

There is abundant evidence for vocalization plasticity in several mammalian species like elephants, bats, seals, and dolphins, not to speak of birds, especially songbirds (Bolhuis et al., 2010). More generally, we may argue that body gestural communication is a widespread characteristic of vertebrates, while vocal communication (innate or learned) has become an important communication pathway only in some lineages. Learned vocalizations are present in even fewer species, coexisting with hand or grasping abilities only in humans, whereas most other vocal learners lack this capacity. Interestingly, cerebral dominance for vocalizations has been reported in many species, both vocal learners and non-vocal learners (Corballis, 2003). Thus, at least in mammals there seems to be no phylogenetic association between grasping abilities and the capacity for vocal learning or imitation. Birds have grasping feet, but it is not known if this ability involves a mirror neuron system, or if its neural representation matches the neural substrate for vocalizations. More likely, imitation tends to be more conspicuous in animals that have developed vocal learning, suggesting that the latter is more closely associated with the acquisition of imitative capacities. Along this line, a vocalization mirror neuron system has been proposed to exist in songbirds, but this possibility and the relation of this putative circuit with a grasping mirror system have yet to be proved (Bonini and Ferrari, 2011).

The grasping mirror neuron network is an ancient characteristic of the primate brain, and therefore cannot by itself account for the origin of vocal language. Among other capacities, an emerging language may have needed shared intentionality, mirror neuron properties, and the capacity to understand actions (Premack, 2004; Tomasello et al., 2005; Corballis, 2010). However, the mirror neuron-gestural perspective does not provide any clue as to how or why speech emerged and became the dominant communication channel. More likely, the key event was the reinforcement of a primitive auditory–vocal sensorimotor circuitry, which, as it expanded, probably took advantage of circuits previously involved in other motor functions, recruiting them for vocal control mechanisms.

TOOL USE, GESTURES, AND A PRIMITIVE SEMANTICS

From the mirror neuron perspective, gestures have been proposed to be crucial for the acquisition of a primitive semantics (Arbib, 2005). In this process, grasping ability and voluntary hand control may have been important elements to facilitate shared attention, and possibly led to the appearance of pointing behavior, which is critical for making reference to the world (Call, 1980). From pointing, other meaningful hand gestures may have evolved, especially in the context of a primitive tool-making and tool-using technology in which the emulation of tool use may have conveyed a ritualized semantics.

There is an extensive literature on tool manufacturing and use in modern humans, early hominids, and non-human primates (Greenfield, 1991; Boesch, 1993; Call and Tomasello, 2007; Ambrose, 2010; Liebal and Call, 2012; Macellini et al., 2012). Observation of tool use produces activation of a sector in the inferior parietal lobe in humans but not in tool-trained monkeys (Peeters et al., 2009). However, the pattern of brain activation during tool use depends on the tool being used. Comparing two different Paleolithic stone tool tasks, one early (Oldowan), and the other from a later period (Acheulean), Stout and Chaminade (2012) reported that both tasks activated the inferior parietal cortex and the ventral premotor cortex, but only the Acheulean task produced activation of the right inferior frontal gyrus (area 45). These authors and others further propose that tool use and manufacture are hierarchically organized and can be described in a nested syntax, comparable to the recursive syntax of language (Stout and Chaminade, 2012). However, and consistent with the present perspective, they indicate that parsing of behavioral sequences during tool manufacture or use may have provided a bridge between instrumental actions and vocal syntax without the need to invoke a separate communicative gestural stage.

Communicative gestures are derived from non-communicative actions like throwing, grasping or tool use, through a process called ontogenetic ritualization, which may become assimilated during phylogeny (Pika et al., 2005). Orangutans and gorillas have been shown to perform specific gestures that imply distinct meanings, being used intentionally and in a frame of contextual flexibility (Genty et al., 2009; Cartmill and Byrne, 2010). There is also evidence that apes usually incorporate objects in their gestures, and that this correlates with the species' use of tools in the wild (Call and Tomasello, 2007; Liebal and Call, 2012). Pantomimes are gestures resembling the actions they represent but are not effective in performing the action. Whereas in non-human primates these pantomimes are simple representations of actions lacking abstraction, in humans they involve an abstract content, accompany symbolic communication, and may support the signer's capacity for problem solving (Cartmill et al., 2012).

The fact that apes can be taught sign language but are unable to master learned vocalizations has been proposed as supporting a gestural origin for human language (Corballis, 2003). Nevertheless, there is a difference between ontogenetic plasticity and capacity for evolutionary change. A rapid selective trend toward increasing vocal plasticity and vocal control is perfectly possible, and is compatible with the evidence of vocal learning in other mammals and in songbirds (Bolhuis et al., 2010).

However, in this scenario there is little insight into how the transition from gestural references to vocal references could be made. In my view, a gestural pantomime may have been accompanied by the use of sounds imitating the referred object; this simultaneity of gesture and vocalization is likely to have been crucial for the establishment of meaning in vocal behavior (see Taglialatela et al., 2011). Furthermore, increasing vocal plasticity may have facilitated vocal imitation of physical or animal sounds, rapidly taking over most symbolic contents. To what extent this primitive semantics was gesture-based or vocalization-based will probably never be known, but it is likely that there were several ways to convey meaning, and more importantly, individuals used whatever means they had available, be they gestures, signs, or other signals, to call attention to relevant events under different circumstances.

MIRROR NEURONS AND WORKING MEMORY

Recently, there has been an important debate as to whether motor functions are essential or not for speech processing, which impinges into the mirror neuron – vocal learning debate. A current interpretation is that the motor system modulates, but does not obligate speech perception (Hickok et al., 2011a,b). However, this modulation may be what is needed to have a better learning capacity, as children with a stronger verbal working memory end up with a larger vocabulary some years later (Baddeley, 2003). In other words, although it may not be necessary for phonological processing, inner speech may protect a perceptual memory trace from interfering processes, helping its maintenance for a longer time (Baddeley, 2003; Marvel and Desmond, 2012).

Furthermore, mirror neurons may eventually prove to be involved in verbal working memory mechanisms. An important component of working memory capacity depends on the close integration between sensory and motor systems, in which audio–vocal mirror neurons may participate, as is perhaps the case in song-learning birds (Bolhuis et al., 2010; Bonini and Ferrari, 2011). The case of conduction aphasia, involving not only a disruption of the white matter as originally considered, but also lesions in the surrounding cortical areas, is characterized by a dysfunction in short-term memory and in imitative capacities (Trortais, 1974; Buchsbaum et al., 2011; Song et al., 2011), which stresses the relation between imitation, sensorimotor integration, and short-term memory. Again, a commonly involved cortical area in conduction aphasia is the posterior planum temporale, i.e., area Spt (Buchsbaum et al., 2011).

SPEECH, BIRDSONG, AND MIRROR NEURONS: DEEP HOMOLOGY?

Finally, some words on studies of vocal learning in songbirds may be worth mentioning here. This has become a rich scientific program in which very different processes, including adult neurogenesis, neural plasticity, gene expression patterns, and even syntactical learning have been addressed (Bolhuis et al., 2010; Abe and Watanabe, 2011; Berwick et al., 2011), confirming Darwin's original speculation of a parallel between speech and birdsong. Moreover, in songbirds, the vocal learning circuit has a similar (but not homologous) architecture as the language circuits, involving cortico-basal ganglia–thalamic circuits (Bolhuis et al., 2010).

In the present context, it may be relevant to mention the recent proposal of a "deep homology" (homology at the gene level) between vocal learning mechanisms in songbirds and humans, based on the participation of the gene FOXP2 in this process (especially in circuits involving the basal ganglia; Scharff and Petri, 2011). FOXP2 is a gene whose mutation causes an inherited verbal dyspraxia in humans, and was initially proposed to be a sort of master-language gene. However, the interpretations of the behavioral phenotype of the affected members are a matter of debate, some proposing that it relates to an inability to denote tense, gender, and other grammatical functions; others view this condition as a phonological articulatory disorder, and still others argue that it affects all levels of language processing (Varga-Khadem et al., 2005). Despite these disagreements, there is evidence that FOXP2 has been a target of selection in the human lineage; it differs from the chimp homolog in two point mutations (Enard et al., 2002; Zhang et al., 2002; Teramitsu et al., 2004; Krause et al., 2007) and is a common transcriptional target of genes displaying accelerated evolution in humans (Lambert et al., 2011). This gene also displays accelerated evolution in echolocating bats, another vocal learning group (Li et al., 2007). Interestingly, in songbirds, FOXP2 expression is modulated during song learning (Haesler et al., 2004; Teramitsu et al., 2010), and its transcript is required for appropriate song learning (Haesler et al., 2007). Furthermore, diminishing FOXP2 expression produces a decrease in dendritic spine density in the basal ganglia song area of the zebra finch (Schulz et al., 2010). However, deficiency of this gene affects the intensity but not the structure of innate vocalizations in mouse pups (Gaub et al., 2010; Fischer and Hammerschmidt, 2011). Furthermore, mutations of FOXP2 produce generalized deficits in synaptic plasticity and motor learning in mice (Groszer et al., 2008). In light of this evidence, FOXP2, rather than a specific language master gene, is now considered to be involved in more general aspects of sensorimotor learning, and may be of particular relevance for the acquisition of complex, learned motor patterns which include birdsong and speech (Varga-Khadem et al., 2005). If this were the case, any FOXP2-dependent process of sensory-guided learning would represent deep homology with the language and the birdsong circuits.

A few years ago, Corballis (2004) suggested a possible link between FOXP2 and the mirror neuron system, based on evidence indicating underactivity in Broca's area in subjects bearing a mutation of this gene (Liégeois et al., 2003; see also Bosman et al., 2004). It is not yet known whether FOXP2 is specifically expressed in hand-grasping processes in non-human primates. If it were, this evidence would be consistent with the above interpretation, namely that FOXP2 underlies a variety of sensorimotor learning processes, including hand-grasping, speech, and birdsong.

CONCLUDING REMARKS

This review mostly uses information on neural connectivity to establish the phylogenetic continuity of neural circuits involved in speech processing. For reasons of space, other aspects like the comparative microanatomy, cross-species volumetric analyses, and the details of behavioral studies have been discussed only briefly.

Summarizing all the information presented, and considering the several discrepancies in some specific issues, I will take the opportunity to make some concluding remarks. First, the cytoarchitectonic homologs to human areas 44 and 45 are the homonymous areas in the monkey. In the latter, area 44 represents an orofacial specialization of the ventral premotor area 6v (F5), receiving inputs from area 45, which conveys facial and auditory information from the anterior temporal lobe. An arcuate fasciculus may be present in the monkey, but it is probably not a robust tract. Inferior parietal areas send projections to the ventral premotor areas and possibly to area 44 of the monkey. There are discrepancies as to the inferoparietal projection to area 45.

It is likely that the dorsal auditory–vocal pathway via the arcuate fasciculus/SLF did not arise out of nothing, and that a rudimentary auditory pathway to the VLPFC strengthened gradually from monkey to chimpanzee to human. In the chimp, these projections may only have a weak participation in vocalization, but in hominids, neighboring inferior parietal areas were recruited to participate in the planning of motor processes involving vocal articulation, using auditory projections carried by the MLF. The ventral pathway became adapted to transmit echoic and semantic information into the anterior Broca's area.

As is possibly the case in songbirds, it is likely that mirror neurons were included in the nascent phonological loop of early humans, an auditory–vocal sensorimotor pathway with sufficient plasticity and memory capacity to learn complex vocal utterances by imitation. Across species, imitative capacity appears to be associated more with vocal learning than with grasping ability. Nonetheless, it is possible that gestures and vocalizations were both initially used to generate shared attention, which may be a requisite for a primitive semantics. The simultaneity of gestures and vocalizations was likely an important element to transmit stronger messages, and as vocalizations became increasingly sophisticated, they became dominant over gestures.

Thus, human communication is, and has always been, multimodal and opportunistic, using whatever means are available to transmit the intended meaning. Indeed our species is characterized by the urge to communicate things (Tomasello et al., 2005). We have developed a specialized neural device, the phonological loop that, together with other cognitive specializations, has propelled our communication capacities far beyond those of other animals. Whenever speech is incapable of transmitting information, we literally use the most handy channel at our disposal. That is why, besides sign language, we have developed writing, which is now being transformed into key-pressing, and may eventually become a fully digitalized system for which we may need minimal motor skills.

ACKNOWLEDGMENTS

This work was supported by the Millenium Center for the Neuroscience of Memory, Chile, NC10-001-F, which is developed with funds from the Innovation for Competitivity from the Ministry for Economics, Fomentation, and Tourism, Chile. I am also grateful to Francisco Zamorano, who prepared the Figures presented in this article; and to John Ewer for his kind help in revising the english language.

REFERENCES

Abe, K., and Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* 14, 1067–1077.

Aboitiz, F. (1995). Working memory networks and the origin of language areas in the human brain. *Med. Hypotheses* 44, 504–506.

Aboitiz, F., Aboitiz, S., and García, R. (2010). The phonological loop: a key innovation in human evolution. *Curr. Anthropol.* 51(Suppl. S1), S55–S65.

Aboitiz, F., and García, R. (1997). The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. *Brain Res. Rev.* 25, 381–396.

Aboitiz, F., and García, R. (2009). Merging of phonological and gestural circuits in early language evolution. *Rev. Neurosci.* 20, 71–84.

Aboitiz, F., García, R., Bosman, C., and Brunetti, E. (2006a). Cortical memory mechanisms and language origins. *Brain Lang.* 98, 40–56.

Aboitiz, F., García, R., Brunetti, E., and Bosman, C. (2006b). "The origin of Broca's area and its connections from an ancestral working/active memory network," in *Broca's Area*, eds K. Amunts and Y. Grodzinsky (Oxford: Oxford University Press), 3–16.

Aboitiz, F., García, R., Brunetti, E., and Bosman, C. (2005). Imitation and memory in language origins. *Neural Netw.* 18, 1357.

Aboitiz, F., López, J., and Montiel, J. (2003). Long distance communication in the human brain: timing constraints for inter-hemispheric synchrony and the origin of brain lateralization. *Biol. Res.* 36, 89–99.

Ambrose, S. H. (2010). Coevolution of composite-tool technology, constructive memory, and language: implications for the evolution of modern human behavior. *Curr. Anthropol.* 51, S135–S147.

Anwander, A., Tittgemeyer, M., Yves von Cramon, D., Friederici, A. D., and Knösche, T. R. (2007). Connectivity-based parcellation of Broca's area. *Cereb. Cortex* 17, 816–825.

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* 28, 105–167.

Arbib, M. A. (2010). Mirror system activity for action and language is embedded in the integrationof dorsal and ventral pathways. *Brain Lang.* 112, 12–24.

Armstrong, A. C., Stokoe, W. C., and Wilcox, S. E. (eds). (1995). *Gesture and the Nature of Language.* Cambridge: Cambridge University Press.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., and Katz, S. (1996). Dissociation of storage and rehearsal in working memory: PET evidence. *Psychol. Sci.* 7, 25–31.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839.

Baddeley, A. D., Papagno, C., and Vallar, G. (1988). When long-term learning depends on short-term memory. *J. Mem. Lang.* 27, 586–595.

Badzakova-Trajkov, G., Häberling, I. S., Roberts, R. P., and Corballis, M. C. (2010). Cerebral asymmetries: complementary and independent processes. *PLoS ONE* 5, e9682. doi:10.1371/journal.pone.0009682

Belin, P., and Zatorre, R. J. (2000). 'What,' 'where' and 'how' in auditory cortex. *Nat. Neurosci.* 3, 965–966.

Belmalih, A., Borra, E., Contini, M., Gerbella, M., Rozzi, S., and Luppino, G. (2009). Multimodal architectonic subdivision of the rostral part (area F5) of the macaque ventral premotor cortex. *J. Comp. Neurol.* 512, 183–217.

Bernal, B., and Altman, N. (2010). The connectivity of the superior longitudinal fasciculus: a tractography DTI study. *Magn. Reson. Imaging* 28, 217–225.

Berwick, R. C., Okanoya, K, Beckers, G. J. L., and Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* 15, 113–122.

Binkofski, F., Buccino, G., Posse, S., Seitz, R. J., Rizzolatti, G., and Freund, H. (1999). A fronto-parietal circuit for object manipulation in man: evidence from an fMRI-study. *Eur. J. Neurosci.* 11, 3276–3286.

Boesch, C. (1993). "Transmission of tool-use in wild chimpanzees," in *Tools, Language and Cognition in Human Evolution*, eds K. R. Gibson and T. Ingold (Cambridge: Cambridge University Press), 171–183.

Bolhuis, J. J., Okanoya, K., and Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* 11, 747–759.

Bonini, L., and Ferrari, P. F. (2011). Evolution of mirror systems: a simple mechanism for complex cognitive functions. *Ann. N. Y. Acad. Sci.* 1225, 166–175.

Bosman, C., García, R., and Aboitiz, F. (2004). FOXP2 and the language working memory system. *Trends Cogn. Sci.* 8, 251–252.

Brownsett, S. L. E., and Wise, R. J. S. (2010). The contribution of the parietal lobes to speaking and writing. *Cereb. Cortex* 20, 517–523.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L, Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., and Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400–404.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C. A., and Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: an fMRI study. *J. Cogn. Neurosci.* 16, 1–14.

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., and Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory – an aggregate analysis of lesion and fMRI data. *Brain Lang.* 119, 119–128.

Buchsbaum, B. R., and D'Esposito, M. (2008). The search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* 20, 762–778.

Buchsbaum, B. R., Olsen, R. K., Koch, P., and Berman, K. F. (2005a). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48, 687–697.

Buchsbaum, B. R., Olsen, R. K., Koch, P. F., Kohn, P., Kippenhan, J. S., and Berman, K. F. (2005b). Reading, hearing, and the planum temporale. *Neuroimage* 24, 444–454.

Bullier, J., Schall, J. D., and Morel, A. (1996). Functional streams in occipito–frontal connections in the monkey. *Behav. Brain Res.* 76, 89–97.

Call, J., and Tomasello, M. (eds). (2007). *The Gestural Communication of Apes and Monkeys.* Mahwah, NJ: Lawrence Erlbaum Associates.

Call, J. D. (1980). Some prelinguistic aspects of language development. *J. Am. Psychoanal. Assoc.* 28, 259–289.

Calvin, W. H. (1983). A stone's throw and its launch window: timing precision and its implications for language and hominid brains. *J. Theor. Biol.* 104, 121–135.

Cartmill, E. A., Beilock, S., and Goldin-Meadow, S. (2012). A word in the hand: action, gesture, and mental representation in humans and non-human primates. *Philos. Trans. R. Soc. B* 367, 129–143.

Cartmill, E. A., and Byrne, R. W. (2010). Semantics of primate gestures: intentional meanings of orangutan gestures. *Anim. Cogn.* 13, 793–804.

Catani, M., and ffytche, D. M. (2005). The rises and falls of disconnection syndromes. *Brain* 128, 2224–2239.

Cavada, C., and Goldman-Rakic, P. (1989). Posterior parietal cortex in rhesus monkey: II. Evidence for segregated cortico-cortical networks linking sensory and limbic areas with the frontal lobe. *J. Comp. Neurol.* 287, 422–445.

Corballis, M. C. (1992). On the evolution of language and generativity. *Cognition* 44, 197–126.

Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language.* Princeton: Princeton University Press.

Corballis, M. C. (2003). From mouth to hand: gesture, speech, and the evolution of right-handedness. *Behav. Brain Sci.* 26, 199–208.

Corballis, M. C. (2004). FOXP2 and the mirror system. *Trends Cogn. Sci.* 8, 95–96.

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain Lang.* 112, 25–35.

Coudé, G., Ferrari, P. F., Rodà, F., Maranesi, M., Borelli, E., Veroni, V., Monti, F., Rozzi, S., and Fogassi, L. (2011). Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS ONE* 6, e26822. doi:10.1371/journal.pone.0026822

de Zubicaray, G., Postle, N., McMahon, K., Meredith, M., and Ashton, R. (2010). Mirror neurons, the representation of word meaning, and the foot of the third left frontal convolution. *Brain Lang.* 112, 77–84.

Deacon, T. W. (1992). Cortical connections of the inferior arcuate sulcus cortex in the macaque brain. *Brain Res.* 573, 8–26.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180.

Dubois, J., Hertz-Pannier, L., Cachia, A., Mangin, J. F., Le Bihan, D., and Dehaene-Lambertz, G. (2009). Structural asymmetries in the infant language and sensorimotor networks. *Cereb. Cortex* 19, 414–423.

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

Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *J. Neurophysiol.* 73, 2608–2611.

Ferrari, P. F., Gallese, V., Rizzolatti, G., and Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur. J. Neurosci.* 17, 1703–1714.

Fiebach, C. J., Schlesewsky, M., and Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of german Wh-questions. *J. Mem. Lang.* 47, 250–272.

Fiebach, C. J., Schlesewsky, M., Lohmann, G., von Cramon, D. Y., and Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: syntactic integration versus syntactic working memory. *Hum. Brain Mapp.* 24, 79–91.

Fischer, J., and Hammerschmidt, K. (2011). Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. *Genes Brain Behav.* 10, 17–27.

Fogassi, L., Gallese, V., Fadiga, L., and Rizzolatti, G. (1998). Neurons responding to the sight of goal directed hand/arm actions in the parietal area PF (7b) of the macaque monkey. *Soc. Neurosci.* 24, abstr. 257.5.

Frey, S., Campbell, J. S. W., Pike, G. B., and Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.* 28, 11435–11444.

Friederici, A. D. (2004). "The neural basis of syntactic processes," in *The Cognitive Neurosciences III*, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 789–801.

Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: syntactic integration versus syntactic working memory. *Hum. Brain Mapp.* 24, 79–91.

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends Cogn. Sci.* 13, 175–181.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., and Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2458–2463.

Fuster, J. M. (1995). *Memory in the Cerebral Cortex.* Cambridge, MA: MIT Press.

Galaburda, A. M., and Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *J. Comp. Neurol.* 190, 597–610.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609.

Gallese, V., Fogassi, L., Fadiga, L., and Rizzolatti, G. (2002). "Action representation and the inferior parietal lobule," in *Attention and Performance XIX. Common Mechanisms in Perception and Action*, eds W. Prinz and B. Hommel (Oxford: Oxford University Press), 247–266.

Gaub, S., Groszer, M., Fisher, S. E., and Ehret, G. (2010). The structure of innate vocalizations in Foxp2-deficient mouse pups. *Genes Brain Behav.* 9, 390–401.

Gentilucci, M., Benuzzi, F., Gangitano, M., and Grimaldi, S. (2001). Grasp with hand and mouth: a kinematic study on healthy subjects. *J. Neurophysiol.* 86, 1685–1699.

Gentilucci, M., and Dalla Volta, R. (2008). Spoken language and arm gestures are controlled by the same motor control system. *Q. J. Exp. Psychol.* 61, 944–957.

Genty, E., Breuer, T., Hobaiter, C., and Byrne, R. W. (2009). Gestural communication of the gorilla (*Gorilla gorilla*): repertoire, intentionality and possible origins. *Anim. Cogn.* 12, 527–546.

Gerardin, E., Sirigu, A., Lehericy, S., Poline, J. B., Gaymard, B., Marsault, C., Agid, Y., and Le Bihan, D. (2000). Partially overlapping neural networks for real and imagined hand movements. *Cereb. Cortex* 10, 1093–1104.

Gerbella, M., Belmalih, A., Borra, E., Rozzi, S., and Luppino, G. (2007). Multimodal architectonic subdivision of the caudal ventrolateral prefrontal cortex of the macaque monkey. *Brain Struct. Funct.* 212, 269–301.

Gerbella, M., Belmalih, A., Borra, E., Rozzi, S., and Luppino, G. (2010). Cortical connections of the macaque caudal ventrolateral prefrontal areas 45A and 45B. *Cereb. Cortex* 20, 141–168.

Gerbella, M., Belmalih, A., Borra, E., Rozzi, S., and Luppino, G. (2011). Cortical connections of the anterior (F5a) subdivision of the macaque ventral premotor area F5. *Brain Struct. Funct.* 216, 43–65.

Gharbawie, O. A., Stepniewska, I., Qi, H., and Kaas, J. H. (2011). Multiple parietal-frontal pathways mediate grasping in macaque monkeys. *J. Neurosci.* 31, 11660–11677.

Glasser, M. F., and Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cereb. Cortex* 18, 2471–2482.

Greenfield, P. M. (1991). Language, tools and brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. *Behav. Brain Sci.* 14, 531–551.

Grèzes, J., Armony, J. L., Rowe, J., and Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage* 18, 928–937.

Groszer, M., Keays, D. A., Deacon, R. M., de Bono, J. P., Prasad-Mulcare, S., Gaub, S., Baum, M. G., French, C. A., Nicod, J., Coventry, J. A., Enard, W., Fray, M., Brown, S. D., Nolan, P. M., Pääbo, S., Channon, K. M., Costa, R. M., Eilers, J., Ehret, G., Rawlins, J. N,,, and Fisher, S. E. (2008). Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr. Biol.* 18, 354–362.

Häberling, I. S., Badzakova-Trajkov, G., and Corballis, M. C. (2011). Callosal tracts and patterns of hemispheric dominance: a combined fMRI and DTI study. *Neuroimage* 54, 779–786.

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., and Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol.* 5, e321. doi:10.1371/journal.pbio.0050321

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D., and Scharff, C. (2004). FoxP2 expression in avian vocal learners and non-learners. *J. Neurosci.* 24, 3164–3375.

Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., and Siebner, H. R. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16494–16499.

Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Curr. Anthropol.* 12, 5–24.

Hickok, G. (2009). The functional neuroanatomy of language. *Phys. Life Rev.* 6, 121–143.

Hickok, G., Houde, J., and Rong, F. (2011a). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422.

Hickok, G., Costanzo, M., Capasso, R., and Miceli, G. (2011b). The role of Broca's area in speech perception: evidence from aphasia revisited. *Brain Lang.* 119, 214–220.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.

Hopkins, W. D., Russell, J. L., and Schaeffer, J. A. (2012). The neural and cognitive correlates of aimed throwing in chimpanzees: a magnetic resonance image and behavioral study on a unique form of social tool use. *Philos. Trans. R. Soc. B* 367, 37–47.

Iacoboni, M., and D'Apretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nat. Rev. Neurosci.* 7, 942–951.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science* 286, 2526–2528.

Ide, A., Dolezal, C., Fernández, M., Labbé, E., Mandujano, R., Montes, S., Segura, P., Verschae, G., Yarmuch, P., and Aboitiz, F. (1999). Hemispheric differences in the variability of fissural patterns in parasylvian and cingulate regions of human brains. *J. Comp. Neurol.* 410, 235–242.

Izumi, A., and Kojima, S. (2004). Matching vocalizations to vocalizing faces in a chimpanzee (*Pan troglodytes*). *Anim. Cogn.* 7, 179–184.

Johnson Frey, S. H., Maloof, F. R., Newman-Norlund, R., Farrer, C., Inati, S., and Grafton, S. T. (2003). Actions or hand-objects interactions? Human inferior frontal cortex and action observation. *Neuron* 39, 1053–1058.

Josse, G., and Tzourio-Mazoyer, N. (2004). Hemispheric specialization for language. *Brain Res. Rev.* 44, 1–12.

Jürgens, U., and Alipour, M. (2002). A comparative study on the corticohypoglossal connections in primates, using biotin dextran amine. *Neurosci. Lett.* 328, 245–248.

Kaas, J. H., Gharbawie, O. A., and Stepniewska, I. (2011). The organization and evolution of dorsal stream multisensory motor pathways in primates. *Front. Neuroanat.* 5:34. doi:10.3389/fnana.2011.00034

Kaas, J. H., and Hackett, T. A. (1999). 'What' and 'where' processing in auditory cortex. *Nat. Neurosci.* 2, 1045–1047.

Keller, S. S., Roberts, N., and Hopkins, W. (2009). A comparative magnetic resonance imaging study of the anatomy, variability, and asymmetry of Broca's area in the human and chimpanzee brain. *J. Neurosci.* 29, 14607–14616.

Kelly, K., Uddin, L. Q., Shehzad, Z., Margulies, D. S., Castellanos, F. X., Milham, M. P., and Petrides, M. (2010). Broca's region: linking human brain functional connectivity data and nonhuman primate tracing anatomy studies. *Eur. J. Neurosci.* 32, 383–398.

Kendon, A. (2004). *Gesture: Visible Action.* Cambridge: Cambridge University Press.

Keysers, C., Kohler, E., Umiltá, M. A., Nanetti, L., Fogassi, L., and Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Exp. Brain Res.* 153, 628–636.

Kilner, K. M. (2011). More than one pathway to action understanding. *Trends Cogn. Sci.* 15, 352–357.

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

Lambert, N., Lambot, M. A., Bilheu, A., Albert, V., Englert, Y., Libert, F., Noel, J. C., Sotiriou, C., Holloway, A. K., Pollard, K. S., Detours, V., and Vanderhaeghen, P. (2011). Genes expressed in specific areas of the human fetal cerebral cortex display distinct patterns of evolution. *PLoS ONE* 6, e17753. doi:10.1371/journal.pone.0017753

Leichnetz, G. R. (2001). Connections of the medial posterior parietal cortex (area 7m) in the monkey. *Anat. Rec.* 263, 215–236.

Li, G., Wang, J., Rossiter, S. J., Jones, G., and Zhang, S. (2007). Accelerated FOXP2 evolution in echolocating bats. *PLoS ONE* 2, e900. doi:10.1371/journal.pone.0000900

Liebal, K., and Call, J. (2012). The origins of non-human primates' manual gestures. *Philos. Trans. R. Soc. Lond. B* 367, 118–128.

Liégeois, F., Baldeweg, T., Connelly, A., Gadian, D. G., Mishkin, M., and Vargha-Khadem, F. (2003). Language fMRI abnormalities associated with FOXP2 gene mutation. *Nat. Neurosci.* 6, 1230–1237.

Lonsdorf, E. V., and Hopkins, W. D. (2005). Wild chimpanzees show population-level handedness for tool use. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12634–12638.

Lopez-Barroso, D., de Diego-Balaguer, R., Cunillera, T., Camara, E., Münte, T. F., and Rodriguez-Fornells, A. (2011). Language learning under working memory constraints correlates with microstructural differences in the ventral language pathway. *Cereb. Cortex* 21, 2742–2750.

Luppino, G., Murata, A., Govoni, P., and Matelli, M. (1999). Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4). *Exp. Brain Res.* 128, 181–187.

Macellini, S., Maranesi, M., Bonini, L., Simone, L., Rozzi, S., Ferrari, P. F., and Fogassi, L. (2012). Individual and social learning processes involved in the acquisition and generalization of tool use in macaques. *Philos. Trans. R. Soc. B* 367, 24–36.

Maeda, F., Kleiner-Fisman, G., and Pascual-Leone, A. (2002). Motor facilitation while observing hand actions: specificity of the effect and role of observer's orientation. *J. Neurophysiol.* 87, 1329–1335.

Marvel, C. L., and Desmond, J. E. (2012). From storage to manipulation: how the neural correlates of verbal working memory reflect varying demands on inner speech. *Brain Lang.* 120, 42–51.

Matelli, M., Camarda, R., Glickstein, M., and Rizzolatti, G. (1986). Afferent and efferent projections of the inferior area 6 in the macaque monkey. *J. Comp. Neurol.* 251, 281–298.

Meister, I. G., Boroojerdi, B., Foltys, H., Sparing, R., Huber, W., and Topper, R. (2003). Motor cortex hand area and speech: implications for the development of language. *Neuropsychologia* 41, 401–406.

Mesulam, M. M., Van Hoesen, G. W., Pandya, D. N., and Geschwind, N. (1977). Limbic and sensory connections of the inferior parietal lobule (area PG) in the rhesus monkey: a study with a new method for horseradish peroxidase histochemistry. *Brain Res.* 136, 393–414.

Molenberghs, P., Cunnington, R., and Mattingley, J. B. (2009). Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neurosci. Biobehav. Rev.* 33, 975–980.

Moser, D., Baker, J. M., Sanchez, C. E., Rorden, C., and Fridriksson, J. (2009). Temporal order processing of syllables in the leftparietal lobe. *J. Neurosci.* 29, 12568–12573.

Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G., and Orban, G. A. (2011). Action observation circuits in the macaque monkey cortex. *J. Neurosci.* 31, 3743–3756.

Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., and Orban, G. A. (2005). Observing others: multiple action representation in the frontal lobe. *Science* 310, 332–336.

Pandya, D. N., and Yeterian, E. H. (1985). "Architecture and connections of cortical association areas," in *Cerebral Cortex*, Vol. 4, eds A. Peters and E. G. Jones (New York: Plenum Press), 3–61.

Parker, G. J. M., Luzzi, S., Alexander, D. C., Wheeler-Kingshott, C. A. M., Ciccarelli, O., and Ralph, M. A. L. (2005). Lateralization of ventral and dorsal auditory language pathways in the human brain. *Neuroimage* 24, 656–666.

Paulesu, E., Frith, C. D., and Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature* 362, 34–25.

Peeters, R., Simone, L., Nelissen, K., Fabbri-Destro, M., Vanduffel, W., Rizzolatti, G., and Orban, G. A. (2009). The representation of tool use in humans and monkeys: common and uniquely human features. *J. Neurosci.* 29, 11523–11539.

Perrett, D. I., Mistlin, A. J., Harries, M H., and Chitty, A. J. (1990). "Understanding the visual appearance and consequence of hand actions," in *Vision and Action: The Control of Grasping*, ed. M. A. Goodale (Norwood, NJ: Ablex), 163–342.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (2008). A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–374.

Petrides, M., Cadoret, G., and Mackey, S. (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature* 435, 1235–1238.

Petrides, M., and Pandya, D. N. (1984). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J. Comp. Neurol.* 228, 105–116.

Petrides, M., and Pandya, D. N. (1988). Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J. Comp. Neurol.* 273, 52–66.

Petrides, M., and Pandya, D. N. (1999). Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *Eur. J. Neurosci.* 11, 1011–1036.

Petrides, M., and Pandya, D. N. (2002). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur. J. Neurosci.* 16, 291–310.

Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7, e1000170. doi:10.1371/journal.pbio.1000170

Pika, S., Liebal, K., Call, J., and Tomasello, M. (2005). Gestural communication of apes. *Gesture* 5, 41–56.

Premack, D. (2004). Is language the key to human intelligence? *Science* 303, 318–320.

Preuss, T. (1995). "The argument from humans to animals in cognitive neuroscience," in *The Cognitive Neurosciences*, ed. M. S. Gazzaniga (Boston: MIT Press), 1227–1241.

Preuss, T., and Goldman-Rakic, P. S. (1991a). Myelo- and cytoarchitecture of the granular frontal cortex and surrounding regions in the strepsirhine primate Galago and the anthropoid primate Macaca. *J. Comp. Neurol.* 310, 429–474.

Preuss, T., and Goldman-Rakic, P. S. (1991b). Architectonics of the parietal and temporal association cortex in the strepsirhine primate Galago compared to the anthropoid primate Macaca. *J. Comp. Neurol.* 310, 475–506.

Preuss, T., and Goldman-Rakic, P. S. (1991c). Ipsilateral cortical connections of granular frontal cortex in the strepsirhine primate Galago, with comparative comments on anthropoid primates. *J. Comp. Neurol.* 310, 507–549.

Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11800–11806.

Reuland, E. (2010). Imagination, planning, and working memory: the emergence of language. *Curr. Anthropol.* 51(Suppl. 1), S99–S110.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., and Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.

Ringo, J. L., Doty, R. W., Demeter, S., and Simard, P. Y. (1994). Time is of the essence: a conjecture that hemispheric specialization arises from inter-hemispheric conduction delay. *Cereb. Cortex* 4, 331–343.

Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194.

Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141.

Rizzolatti, G., and Luppino, G. (2001). The cortical motor system. *Neuron* 31, 889–901.

Rizzolatti, G. W., and Craighero, L. (2004). The mirror-neuron system. *Ann. Rev. Neurosci.* 27, 169–192.

Rolheiser, T., Stamatakis, E. A., and Tyler, L. K. (2011). Dynamic processing in the human language system: synergy between the arcuate fasciculus and extreme capsule. *J. Neurosci.* 31, 16949–16957.

Romanski, L. M. (2007). Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cereb. Cortex* 17(Suppl. 1), i61–i69.

Romanski, L. M., Averbeck, B. B., and Diltz, M. (2005). Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J. Neurophysiol.* 93, 734–747.

Romanski, L. M., Bates, J. F., and Goldman-Rakic, P. S. (1999a). Auditory belt and parabelt projections to the prefrontal cortex in the rhesusmonkey. *J. Comp. Neurol.* 403, 141–157.

Romanski, L. M., Tian, B., Mishkin, M., Goldman-Rakic, P. S., and Raushecker, J. P. (1999b). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136.

Romanski, L. M., and Goldman-Rakic, P. S. (2002). An auditory domain inprimate prefrontal cortex. *Nat. Neurosci.* 5, 15–16.

Saffran, E. M., and Marin, O. S. (1975). Immediate memory for word lists and sentences in a patient with deficient auditory short-term memory. *Brain Lang.* 2, 420–433.

Salmon, E., Van der Linden, M., Collette, F., Delfiore, G., Maquet, P., Degueldre, C., Luxen, A., and Franck, G. (1996). Regional brain activity during working memory tasks. *Brain* 119, 1617–1625.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., and Weiller, C. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040.

Scharff, C., and Petri, J. (2011). Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 2124–2140.

Schenker, N. M., Buxhoeveden, D. P., Blackmon, W. L., Amunts, K., Zilles, K., and Semendeferi, K. (2008). A comparative quantitative analysis of cytoarchitecture and minicolumnar organization in Broca's area in humans and great apes. *J. Comp. Neurol.* 510, 117–128.

Schulz, S. B., Haesler, S., Scharff, C., and Rochefort, C. (2010). Knockdown of FOXP2 alters spine density in Area X of the zebra finch. *Genes Brain Behav.* 9, 732–740.

Seltzer, B., and Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res.* 149, 1–24.

Seltzer, B., and Pandya, D. N. (1994). Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *J. Comp. Neurol.* 343, 445–463.

Shallice, T., and Warrington, E. K. (1970). Independent functioning of verbal memory stores: a neuropsychological study. *Q. J. Exp. Psychol.* 22, 261–273.

Sherwood, C. C., Broadfield, D. C., Holloway, R. L., Gannon, P. J., and Hof, P. R. (2003). Variability of Broca's area homologue in African great apes: implications for language evolution. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 271, 276–285.

Smith, E. E., and Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proc. Natl. Acad. Sci. U.S.A.* 95, 12061–12068.

Song, X., Dornbos, D. III., Lai, Z., Zhang, Y., Li, T., Chen, H., and Yang, Z. (2011). Diffusion tensor imaging and diffusion tensor imaging-fibre tractograph depict the mechanisms of Broca-like and Wernicke-like conduction aphasia. *Neurol. Res.* 33, 529–535.

Steele, J., and Uomini, N. (2009). Can the archaeology of manual specialization tell us anything about language evolution? A survey of the state of play. *Camb. Archaeol. J.* 19, 97–110.

Stout, D., and Chaminade, T. (2012). Stone tools, language and the brain in human evolution. *Philos. Trans. R. Soc. Lond. B* 367, 75–87.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., and Hopkins, W. D. (2008). Communicative signaling activates "Broca's" homolog in chimpanzees. *Curr. Biol.* 18, 343–348.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., and Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PLoS ONE* 6, e18852. doi: 10.1371/journal.pone.0018852

Teramitsu, I., Kudo, L. C., London, S. E., Geschwind, D. H., and White, S. A. (2004). Parallel FOXP1 and FOXP2 expression in songbird and human brain predicts functional interaction. *J. Neurosci.* 24, 3152–3163.

Teramitsu, I., Poopatanapong, A., Torrisi, S., and White, S. A. (2010). Striatal FOXP2 is actively regulated during songbird sensorimotor learning. *PLoS ONE* 5, e8548. doi:10.1371/journal.pone.0008548

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–293.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691; discussion 691–735.

Trortais, A. (1974). Impairment of memory for sequences in conduction aphasia. *Neuropsychologia* 12, 355–366.

Turkeltaub, P. E., and Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain Lang.* 114, 1–15.

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231–270.

Varga-Khadem, F., Gadian, D. G., Copp, A., and Mishkin, M. (2005). FOXP2 and the neuroanatomy of speech and language. *Nat. Rev. Neurosci.* 6, 131–138.

Vogt, S., Buccino, G., Wohlschläger, A. M., Canessa, N., Shah, N. J., Zilles, K., Eickhoff, S. B., Freund, H. J., Rizzolatti, G., and Fink, G. R. (2007). Prefrontal involvement in imitation learning of hand actions: effects of practice and expertise. *Neuroimage* 37, 1371–1383.

Willems, R. M., Ozyürek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17, 2322–2333.

Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., Ogar, J. M., DeLeon, J., Miller, B. L., and Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron* 72, 397–403.

Yeterian, E. H., Pandya, D. N., Tomaiuolo, F., and Petrides, M. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48, 58–81.

Zhang, J., Webb, D. M., and Podlaha, O. (2002). Accelerated protein evolution and origins of human-specific features: FOXP2 as an example. *Genetics* 162, 1825–1835.

Conflict of Interest Statement: The author declares 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: 28 October 2011; paper pending published: 06 December 2011; accepted: 11 January 2012; published online: 01 February 2012.

Citation: Aboitiz F (2012) Gestures, vocalizations, and memory in language origins. Front. Evol. Neurosci. ***4****:2. doi: 10.3389/fnevo.2012.00002*

frontiers in
EVOLUTIONARY NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 23 December 2011
doi: 10.3389/fnevo.2011.00009

The evolution of syntax: an exaptationist perspective

W. Tecumseh Fitch*

Department of Cognitive Biology, University of Vienna, Vienna, Austria

Edited by:
Constance Scharff, Freie Universitaet Berlin, Germany

Reviewed by:
David W. Dickins, University of Liverpool, UK
Paul M. Nealen, Indiana University of PA, USA

****Correspondence:***
W. Tecumseh Fitch, Department of Cognitive Biology, 14 Althanstrasse, Vienna, Austria.
e-mail: tecumseh.fitch@univie.ac.at

The evolution of language required elaboration of a number of independent mechanisms in the hominin lineage, including systems involved in signaling, semantics, and syntax. Two perspectives on the evolution of syntax can be contrasted. The "continuist" perspective seeks the evolutionary roots of complex human syntax in simpler combinatory systems used in animal communication systems, such as iteration and sequencing. The "exaptationist" perspective posits evolutionary change of function, so that systems today used for linguistic communication might previously have served quite different functions in earlier hominids. I argue that abundant biological evidence supports an exaptationist perspective, in general, and that it must be taken seriously when considering language evolution. When applied to syntax, this suggests that core computational components used today in language could have originally served non-linguistic functions such as motor control, non-verbal thought, or spatial reasoning. I outline three specific exaptationist hypotheses for spoken language. These three hypotheses each posit a change of functionality in a precursor circuit, and its transformation into a neural circuit or region specifically involved in language today. Hypothesis 1 suggests that the precursor mechanism for intentional vocal control, specifically direct cortical control over the larynx, was manual motor control subserved by the cortico-spinal tract. The second is that the arcuate fasciculus, which today connects syntactic and lexical regions, had its origin in intracortical connections subserving vocal imitation. The third is that the specialized components of Broca's area, specifically BA 45, had their origins in non-linguistic motor control, and specifically hierarchical planning of action. I conclude by illustrating the importance of both homology (studied via primates) and convergence (typically analyzed in birds) for testing such evolutionary hypotheses.

Keywords: syntax, cognitive biology, evolution of language, comparative neuroscience, arcuate fasciculus, vocal imitation, human evolution, Broca's area

INTRODUCTION

NATURALIZING THE LINGUISTIC BRAIN

Human language, in its mature state, is a complex system that allows us to encode and communicate thoughts and experiences via hierarchically structured signals called sentences. Humans, apparently uniquely among living animals, are born with a capacity to acquire such systems readily and without the need for explicit tutelage: an instinct to learn language. The capacity for syntax is a central component of this system.

The traditional textbook picture of the neural circuitry underlying the human capacity for language involves premotor speech and syntax areas in Broca's region, posterior temporo-parietal areas involved in comprehension, and connections between them involving the arcuate fasciculus. Current opinions about this three-component model vary from tacit acceptance in many studies to flat-out rejection by a few (e.g., Lieberman, 2000). Today, it seems clear from modern brain imaging work that parts of this model were right all along (e.g., that Broca's area, especially Brodmann's Area (BA) 44 and BA 45, play an important role in language). Other parts are only partially correct (e.g., brain regions involved in language, especially production, tend to be biased toward the left hemisphere, but earlier scientists erred in seeing this bias as a complete left localization). The "connectionist" (in the Lichtheimian nineteenth century sense, not the twentieth century neural network sense) component of the model has had new life breathed into it by comparative diffusion-tensor imaging (DTI) research (Catani and Mesulam, 2008; Rilling et al., 2008). While the importance of left-lateralization, or the precision with which "Wernicke's area" can be localized, may be questioned, the basic notion that Broca's area and its connections with parietal and temporal cortex play an important role in language, both in speech and in syntax, seems rather clear. In this article I will take this for granted, and inquire into the evolution of these linguistic circuits.

When we inquire into the evolutionary changes characterizing human language, our most solid point of reference will always be *differences* between human brains and brains of other primates, whether in brain anatomy or in brain function. That is, in evolutionary terms, we are seeking human *autapomorphies* (traits which differentiate us from our closest relatives, and were not present in the last shared ancestor). Cladistically speaking, the correct comparisons are with our nearest living cousins, the great apes, though both technical and ethical problems with ape neuroscience often force us to rely heavily on more distant primate cousins, especially macaques, to identify evaluate human specializations. So, given the evident absence of language in apes, the fundamental question is "what derived features underlying language arose in the human lineage after it separated from our last common ancestor (LCA) with chimpanzees?" This search for differences, for derived traits, is thus a search for human autapomorphies.

However, in asking this question about human specializations we obviously should not ignore the vast amount of neurobiology that we share with other primates, or other vertebrates more generally. Modern neuroscience is full of neural *synapomorphies* (traits we share with relatives, by common descent). With other vertebrates we share virtually all aspects of neurotransmitter chemistry, neuronal morphology, and brainstem circuitry, along with many detailed aspects of neural processing (e.g., the circuitry of the retina or superior colliculus, or the layout of motor circuitry in the cerebellum or basal ganglia). With other mammals we share the possession of a six-layered neocortex and the strongly conserved pattern of connectivity between its layers. When it comes to chimpanzees, it is no exaggeration to say that virtually *all* known aspects of neuroanatomy appear to be shared, aside from gross size differences, and there are actually few known differences between apes and other Old World primates (for an exception see Nimchinsky et al., 1999). So whatever the neural differences underlying the human capacity for language are, they must always be viewed in the context of an ocean of similarity. Human neural autapomorphies are needles in a synapomorphic haystack.

From a biologist's viewpoint, there is of course nothing special about being "special." Every living species possesses autapomorphies of some sort (otherwise it would not be recognizable as a species), and these specialities always are understood in the context of a much larger set of synapomorphies. We should be no more surprised by human neural autapomorphies tied to language than we are to find that star-nosed moles have a highly specialized somatosensory cortex for their nasal "fingers" (Catania, 1999), or that echolocating bats have specialized auditory regions for their echolocation calls (Suga et al., 1987). It is thus unfortunate that when considering our own specializations, humans have a tendency to invest them with especial significance. This is a mistake. Biologists and neuroscientists should not allow our investigation of neural autapomorphies to be conflated with philosophers' long-running search for defining features of humanity, features that, as a modern replacement to the uniquely human soul, might be misused to justify moral or ethical stances. An ethics built upon empirically determined autapomorphies is built upon shifting sands, as scientific findings might demand change at any point. Thus it is important to divest our investigations of any neural differences that exist from larger, and essentially philosophical, questions about why humans are "special" in ethical or moral senses.

Our task, as neurobiologists, is simply to identify the various neural mechanisms underlying language and, using the comparative approach, to determine which are synapomorphic and which are not. And, given that language relies on widely shared neural foundations, we should not be surprised to find abundant synapomorphies, nor should we take such findings as evidence *against* the existence of autapomorphies. My goal in this paper is to explore the evolution of human neural autapomorphies involved in language, with a focus on syntax. This exploration will adopt a fully comparative approach, using what is known from research on non-human primates, mammals more generally, and birds, to contextualize whatever few autapomorphies humans possess in terms of the much wider foundation of shared synapomorphies.

Furthermore, I will attempt to clarify and defend one particular approach to the evolution of novelties, based on the idea of *exaptation*: that complex structures, evolved in one functional context, can change their function, and be put to work in new domains, often carrying with them traces or constraints due to their prior function. This exaptationist perspective is in full accordance with the truism that evolution has no foresight, or with the evocative image of evolution as a "tinkerer," that gets by with whatever parts or raw materials happen to be available (Jacob, 1977, 1982). Nonetheless, the notion of exaptation when applied to human cognitive evolution has evoked skepticism or even hostility (e.g., Dennett, 1995; Buss et al., 1998; Newmeyer, 1998; Pinker and Jackendoff, 2005), even by those who accept the more general notion of evolution as tinkerer. Thus I will start with a brief exploration of the general concept of change of function (which was a cornerstone of Darwin's model of evolution) and its resurrection in the term "exaptation" by Gould and Vrba, and then very briefly outline some of the objections that have been made to this concept, and the closely related concept of "spandrels," as applied to human cognition and language.

Next, I will outline in detail three exaptive hypothesis concerning human language, which illustrate how an explicitly exaptive model, built on a firm comparative foundation, can both make sense of human specializations and provide further testable predictions for neuroscientists "on the front line" of comparative and functional neuroscience. These hypotheses are, for the most part, built upon the ideas of other scholars, though they are not always framed in the context of exaptation. Although I find each of these hypotheses plausible, my goal here is not to convince anyone that they are true. Rather, my purpose is to demonstrate that exaptive hypotheses about language evolution are plausible, and are part and parcel of a standard biological approach. Further, I detail some testable predictions in the hope that some future neuroscientists will attempt to test such models. In the final part of the paper, I will briefly describe the virtues of a broad comparative approach, incorporating research on both mammals and primates (to discern homologies and ancestral states) and birds (to exploit the hypothesis-testing power of convergent evolution). I believe that there is real value to be gained for evolutionary neuroscience in adopting the rigorous and explicit comparative approach that evolutionary biologists have perfected in the last few decades (complete with the admittedly somewhat ungainly terminology of synapomorphies and autapomorphies).

Naturalizing the human brain demands that we acknowledge both its widely shared foundations with other animals, and recognize at the same time its unique specializations. Our understanding of ourselves will be incomplete without a clear-eyed acceptance of both classes of traits. My goal in this paper will be to investigate how autapomorphies characteristic of our species, and specifically involved in language, might have evolved. More specifically, I aim to defend the proposition that change in function, one of Darwin's favorite evolutionary phenomena, might have played an important role in language evolution and particularly for syntax. From this "exaptationist" perspective, I will argue with many others that precursors for some language functions will be found not in primate communication, but in other aspects of behavior such as motor control or thought. This is by no means a new suggestion, but I will attempt to clarify this exaptive perspective by offering three quite specific models of possible exaptive phenomena, where a circuit built for one function is put to use in another, that are tied to human language.

EXAPTATION AND CHANGE OF FUNCTION

Darwin was famously a gradualist, suggesting that his entire theory would fall apart if one example of evolutionary leaps, or saltations, could be shown. Gradualism, at its simplest, is the assumption that the distribution of phenotypes across evolutionary time is smooth and continuous (like body mass or height) rather than discrete and discontinuous (like eye or flower color). This is a commitment to a continuum of *phenotypes*. Darwin's gradualist perspective was weakened considerably in the modern evolutionary synthesis of genetics and evolutionary theory, initially by the fact that genes are, by their nature, discrete. More recently, the assumption of a smooth relation between genotypic and phenotypic change has been challenged by the discovery of homeotic "master control genes" like HOX or PAX genes in which small genetic changes have drastic phenotypic effects. Nonetheless, a smooth continuum of phenotypes appears to be common in many traits, and such smoothness is assumed to be operative by many models of selection (e.g., the Price equation). So gradualism survives, but in a weaker form than Darwin considered to necessary.

A different kind of potential "smoothness" in evolution involves continuity of *function*. In this case, Darwin was a discontinuist, for he gave multiple examples of "changes in function," and he saw these as exemplifying a major driving force in the evolution of complexity. For example, in discussing the evolution of vertebrate lungs, Darwin proposed that these structures first evolved as complex swim bladders involved in buoyancy control in ancestral fish. As our aquatic ancestors became air breathers, the function of these ancestral swim bladders was converted to that of lungs, which function in the respiratory exchange of gases. Darwin cited several other examples of change in function, including the separate use of the swim-bladder in hearing in some fish, and from flowers and barnacles. Although Darwin gave no specific term to "nascent" traits in which function changed, later authors termed the precursor versions "pre-adaptations." More recently (Gould and Vrba, 1982) objected to the term "pre-adaptation" because of its teleological connotations of evolutionary foresight, and coined the term "exaptation" for the end-product of an evolutionary change in function. Gould and Vrba argued convincingly that such changes in function are ubiquitous in evolution, so much that most complex traits have gone through at least one change in function.

Darwin recognized the exaptive process of evolutionary change of function, combined with gradualism, as a crucial factor in the evolution of complex organs for a simple reason: it is often difficult to see how a small rudiment of an organ could have any adaptive function until it exceeded a certain size or complexity. What good is a tiny lung, that could only fulfill a small proportion of the individual's respiratory needs? What good is a tiny wing, that could never allow an organism to become airborne (Kingsolver and Koehl, 1985)? By the assumption of continuity of function, these are big problems. But as soon as we consider the possibility of change of function, the answer becomes obvious: lungs started out as swim bladders (where a little buoyancy might still be useful) and wings as temperature regulators (where a small radiating or insulating surface might make a considerable contribution). For further examples of change of function (see Gould and Vrba, 1982), and for a critique and clarification of some of the conceptual issues this raises (see Reeve and Sherman, 1993; Dennett, 1995). It should be noted that it appears that Darwin was incorrect in his swim-bladder example about the *direction* of change. It now appears that lungs evolved first (in early fish, who also possessed gills) and were later converted to complex swim bladders in later fish radiations [see Fitch (2010) for this and other examples of exaptations related to speech].

EXAPTATION IN COGNITIVE EVOLUTION

Given the established fact that organs can change their function, and the importance of this fact for evolutionary theory, it is somewhat surprising that exaptation has been dismissed (or worse, ignored) by many researchers interested in human cognitive evolution (cf. Gould, 1991; Buss et al., 1998). This is particularly true in language evolution, where a continuity of a function in communication is assumed by many (e.g., Pinker and Bloom, 1990; Pinker and Jackendoff, 2005) while changes in function are seen by others as an explanatory necessity (e.g., Berwick, 1997; Chomsky, 2010).

Part of this debate can be readily resolved by adopting a multi-component perspective, wherein human language capabilities enlist several interacting but independent mechanisms. Many of these mechanisms will be shared with other animals, but several appear to have changed considerably since the LCA of humans and chimpanzees (hereafter, the LCA), including those involved in signaling, syntax, and semantics (cf. Fitch, 2010). From a multi-component perspective it is perfectly plausible that some novel aspects of language (e.g., speech production or hearing) are fully continuous in function from the LCA, whereas others may constitute exaptations where function has changed. Prominent candidates for exaptive change are the computational mechanisms subserving linguistic syntax, the topic of the current article. But to illustrate the point I will start with a less controversial and better understood example, drawn from the neural circuitry underlying language. First it is necessary to clarify how a multi-component perspective on contemporary language circuits leads to implications for a multi-stage model of language evolution.

EXAPTATION AND DIFFERING MODELS OF PROTOLANGUAGE

A logical concomitant of a multi-component perspective on language evolution is a notion of intermediate stages during human evolution in which certain components of language were present, while others were either lacking entirely or not integrated into a linguistic system. For those novel aspects of language that evolved since the LCA, the term "protolanguage" is often used to denote such a precursor system, an evolutionary way-station on the path to full modern language. The term "protolanguage" was first used in an evolutionary context by anthropologist Gordon Hewes (1973), who introduced it in the context of a gestural protolanguage hypothesis, by which early stages of linguistic signaling were in the manual/visual domain. It was further popularized by linguist

Derek Bickerton, who used the term to denote a lexical model of protolanguage, in which spoken words formed the first stage of language evolution (Bickerton, 1990, 2000, 2007). One of the oldest models of protolanguage dates back to Darwin (1871), who suggested that the earliest stages of human language were musical, with no semantic content. This notion of "musical protolanguage" or "prosodic protolanguage" has more recently been embraced by a number of scholars (Richman, 1993; Brown, 2000; Mithen, 2005; Fitch, 2006). For more detail on these and other conceptions of protolanguage (see Fitch, 2010).

Different models of protolanguage constitute different hypotheses about the nature of the mechanisms involved in modern language, along with an implied order in which these mechanisms were acquired during human evolution. Often, these models also offer, whether explicitly or implicitly, hypotheses about the function(s) subserved by different mechanisms as well. For example, Bickerton (2010) suggests that the first words served the communicative function of displaced reference (allowing early hominins to communicate about distant food sources), and that this putative function has remained continuous since that time. In contrast, Darwin's model suggests that the initial function of complex, learned signals was courtship and territorial defense, in direct analogy to birdsong. Darwin's musical protolanguage model thus entails a change of function, from non-verbal communication to semantic communication of propositional knowledge, during hominin evolution. Bickerton's model is thus continuist in this respect, while Darwin's is exaptive, involving a change in function.

Peter MacNeilage, a speech scientist, has offered an exaptive hypothesis for the evolution of speech, where the oscillatory lip, jaw, and tongue movements involved in producing consonant/vowel alternations in human speech we co-opted from the basic mammalian circuitry controlling chewing, licking, and suckling (MacNeilage, 1998). Like Darwin's swim bladder example, MacNeilage's hypothesis suggests that the complex and coordinated motor activities used in speech were taken over from the already complex and coordinated motor control that had evolved, over tens of millions of years, in ancestral mammals to subserve feeding. In the commentary following this article, many critics suggested that a continuity of function was more likely (and thus that human speech evolved from primate vocalizations), but none of them questioned the general plausibility of MacNeilage's exaptive explanation.

Another prominent exaptive model for language evolution is due to Noam Chomsky (1968, 1976, 2010), who argued that the roots of the generative capacity of language should be sought in private conceptual abilities, rather than in communication of knowledge between individuals. His argument, echoed by various prominent biologists like Jacques Monod and Salvador Luria, is that language is far more powerful than it needs to be for communication. Thus, the vast reach of language into all aspects of human cognition cannot be explained by a step-by-step expansion of communication abilities, because most of the sentences we can generate and understand would have no conceivable selective advantage (cf. Piattelli-Palmarini, 1994; Berwick, 1997). In contrast, a flexible and extendable system of thought would always be of considerable adaptive value, because even small advantages in thought can prove valuable if they allow an individual to out-think conspecifics (cf. Jolly, 1966; Humphrey, 1976; Byrne, 1997). In other words, language appears to be more powerful than needed for everyday communication but appropriately powerful as a medium for creative, generative thought. Again, this is simply an exaptive hypothesis like many others, where the function of one aspect of language has changed, over evolutionary time, from conceptual to communicative, but this particular exaptive hypothesis has been singled out for criticism by many authors (Pinker and Bloom, 1990; Newmeyer, 1998; Jackendoff and Pinker, 2005; Pinker and Jackendoff, 2005), so much so that linguist Frederick Newmeyer confessed to finding Chomsky's ideas on this front "utterly implausible" (Newmeyer, 1998).

As this brief review illustrates, there is considerable disagreement about the importance of exaptive change in function in the evolution of human language, and opinions are particularly polarized in the domain of syntax. Of course, one's model of language evolution depends on one's conception of contemporary language, and some real disagreements stem from such underlying theoretical biases. On the other hand, the many examples of change of function in other evolutionary domains force us to consider exaptive hypotheses concerning language with an open mind. The conversion between swim bladders and lungs, or between radiators and insect wings, may not appear intuitively plausible, but both are taken very seriously by contemporary evolutionary biologists.

Furthermore, when continuity of function is simply assumed without argument, this assumption may conceal further problematic assumptions. For example, Clark (2011) shows how Bickerton's model of language evolution, based on continuity of communicative function, makes several problematic implicit assumptions that Bickerton himself fails to consider. Furthermore, the assumption of functional continuity may block consideration of alternative hypotheses. For example, it was assumed for many years that the *only* plausible explanation for the descent of the human larynx was for speech communication (Lieberman, 1968, 1975, 1984). This led researchers to overlook examples of laryngeal descent in multiple animal species (e.g., Fitch and Reby, 2001; Weissengruber et al., 2002; Frey and Riede, 2003), all of which lack speech, and which clearly demonstrate that a descended larynx *can* serve adaptive functions other than spoken language (see below for the importance of convergent evolution in such inferences).

Thus, rather than formulating the problem of language evolution as the transition from non-human primates to modern human language, in all its glory, we should rather investigate models in which several stages of protolanguage existed. It is less important to know *when* these different stages occurred (e.g., in *Australopithecus* or *Homo erectus*) than it is to construct plausible scenarios in which specific novel abilities evolved, building on pre-existing mechanisms but extending them in the required direction. For example, we know that humans evolved novel speech control mechanisms, so one stage in human evolution must certainly have involved this capacity. However, many organisms have vocal imitation but lack language, so we must further posit additional evolutionary events in which syntactic or semantic abilities evolved. It is of course possible that multiple capacities evolved simultaneously, but to me it seems much more likely that each of these capacities evolved at different times, involved different circuits, and possibly served different adaptive functions. It is equally possible that some of these stages were exaptive (building on non-communicative precursors) while others were characterized by continuity of function.

THREE EXAPTIVE HYPOTHESES IN THE EVOLUTION OF LANGUAGE

In the rest of this article I will consider the evolution of the various mechanisms underlying modern linguistic abilities, taking seriously the possibility of one or more changes in function.

EXAPTIVE HYPOTHESIS 1: SPEECH MOTOR CONTROL AND DIRECT CORTICO-AMBIGUAL CONNECTIONS

While humans are gifted vocal imitators, picking up words and melodies in their environment spontaneously, other primates are not: a chimpanzee raised in a human home will never learn to speak or sing (Yerkes and Yerkes, 1929; Hayes, 1951). This is not due to their peripheral vocal anatomy: most mammals can dynamically lower their larynx to a position that would allow different vowels to be produced (Fitch, 2002, 2010). Furthermore, vocal tract deficiencies would not account for the failure of chimpanzees to imitate melodies, which would require only laryngeal control. Thus, the biological explanation of the lack of complex vocal learning in non-human primates must have something to do with brain structure.

A prominent hypothesis for what, precisely, accounts for the difference between humans and other primates was developed by Uwe Jürgens and his colleagues, on the basis of initial observations by the comparative neuroanatomist Hans Kuypers, and I thus call it the "Kuypers/Jürgens direct connections hypothesis" (Kuypers, 1958a,b, 1973; Jürgens et al., 1982). Kuypers explored the connections from motor cortex to subcortical motor systems in the brainstem and spinal cord, and compared these between cats, non-human primates (chimpanzees and macaques), and humans. He used Nauta/Gygax staining to document axonal degeneration after lesions (in animals) or strokes (in humans) to the motor cortex. He examined in particular the motor neurons controlling laryngeal muscles, located in the nucleus ambiguus of the medulla. He found that there were projections from cortical motor neurons directly onto these motor neurons only in humans. In cats or non-human primates, only multi-synaptic indirect connections were present to ambigual motor neurons. Interestingly, Kuypers found that chimpanzees and macaques *do* have direct connections to brainstem nuclei controlling the face (lips and jaw), while cats lack these.

This argument is consistent with lesion data: while lesions to motor cortex can induce long-lasting mutism in humans, matched lesions have no effect upon vocal production in monkeys (Sutton et al., 1974; Aitken, 1981; Jürgens et al., 1982). Lesions to lateral cortex often severely disrupt human speech, but spare innate species-typical vocalizations like laughter and crying (Foerster, 1936; Groswasser et al., 1988). Finally, since non-human primates *have* direct cortico-motor connections to the jaws, tongue, and lips, the frequent use by chimpanzees of learned, un-phonated "vocalizations" like jaw clacks, lip smacks, and lip buzzes ("raspberries") in volitional, goal-directed communication (Marshall et al., 1999; Reynolds Losin et al., 2008) is consistent with the general idea that direct connections are needed for learned, volitional communication. All that is missing in chimpanzees is laryngeal control. These data, and other convergent data from birds (see below), has led many comparative neuroscientists to endorse the idea that direct cortico-motor connections to the larynx play a key role in human speech abilities (e.g., Myers, 1976; Deacon, 1992; Holstege and Ehling, 1996; Striedter, 2004).

How did humans develop direct connections that are lacking in other primates? In this case we have a very clear candidate for a mammalian "pre-adaptation" for this human situation: the cortico-spinal tract. The cortico-spinal tract is a major descending pathway from motor cortex down to motor neurons within the spinal cord. Cortico-spinal axons originate in pyramidal neurons in layer V of the neocortex, mostly in primary motor cortex but also from premotor cortex, the supplementary motor area (SMA), cingulate gyrus, and somatosensory cortex. A clear homolog of the cortico-spinal tract is present in all mammals. A closely related set of cortico-motor projections make up the cortico-bulbar tract, which project from cortex down to various brainstem motor nuclei found in the trigeminal (CN V, controlling the jaw), facial (CN VII: lips and other facial musculature), and hypoglossal (CN XII, controlling the tongue) nuclei. Thus, in primates, the cortico-bulbar and cortico-spinal tracts together innervate motor neurons above and below the key laryngeal motor neurons located in the nucleus ambiguus (CN X, containing the motor neurons of the vagus nerve complex, including the superior and recurrent laryngeal nerves). I suggest that either, or both, of these tracts provided a precursor of the direct cortico-ambigual connections found in humans. Specifically, I propose that the developmental processes underlying axonal growth and synapse-formation in the cortico-spinal tract were "exapted" to generate cortico-ambigual connections found in adult humans and which play a key role in speech.

Why do such direct connections develop in humans but not other primates? An intriguing hypothesis has been put forward by Terrence Deacon, involving competition between an innate call system based in the brainstem, and the cortico-motor system. Most mammals share a basic system for producing species-typical calls, involving a "control center" in the periaqueductal gray (PAG) of the midbrain (Jürgens, 1994; Holstege and Ehling, 1996). This system can produce the entire species-typical vocal repertoire of cats or squirrel monkeys, or innate vocalizations like laughter and crying in humans, and does not rely upon cortical input (hence, anencephalic human babies who lack a neocortex can still smile, laugh, and cry normally). Deacon has suggested that, during development, there is competition between this prepotent brainstem system and cortical axons projecting down through the brainstem, which transiently connect to laryngeal motor neurons in the nucleus ambiguus, but are out-competed by the innate vocal system projections from PAG and other centers (Deacon, 1992, 1997). By Deacon's hypothesis, evolving the kind of direct cortical control over vocalizations typical of humans would *require* a reduction in the innate call system, and perhaps also the innate call repertoire. This hypothesis has various testable predictions, and even suggests that direct cortico-motor connections could be elicited experimentally in other mammals, if the innate call system was reduced by lesions at the right stage of development (along the lines of Roe et al., 1992).

Obviously, speech motor control is closely tied to the auditory/vocal modality of communication, and speech is not synonymous with, or even necessary for, human language. However, this aspect of speech does provide a simple, clear example of how a particular neural circuit involved in spoken language might have evolved via exaptation of a pre-existing motor circuit found in other primates. While the Kuypers/Jürgens hypothesis is the best available explanation of human vocal control capacities at present, the exaptive hypothesis sketched above remains speculative at present. However, it is clearly both testable, and consistent with a considerable body of knowledge from comparative neuroscience.

EXAPTIVE HYPOTHESIS 2: INTRACORTICAL CONNECTIONS INVOLVED IN SPEECH

While the previous example focused on speech motor output, another well-known example of an important circuit involved in spoken language is provided by the *arcuate fasciculus* connecting temporal/parietal regions to motor and premotor cortex. This white-matter fiber tract was originally discussed in Wernicke's analysis of "conduction aphasia" (Wernicke, 1874), but its importance was highlighted in some classic papers by Norman Geschwind (1965, 1970). Wernicke's original argument concerned the necessity, in speech, of an exchange of information between auditory areas which store the sound of a word and motor areas which mediate the motor outputs capable of generating that sound with the articulators. This idea was (over)extended by Lichtheim and Geschwind to include a third hypothetical association area in the angular gyrus. But Wernicke's basic notion that the ability to learn to vocally produce heard sounds (such as words or melodies) requires connections between auditory and motor centers is sound, and has withstood the test of time. For a detailed history of this discussion (see Catani and Mesulam, 2008).

The advent of DTI and *in vivo* tractography has breathed considerable new life into these old ideas, documenting differences between human brains and those of other primates (Rilling et al., 2008). As in the previous example, there appear to be significant differences among species in the intracortical white-matter connections between temporal, parietal, and frontal regions that are plausibly related to speech motor control (see Petkov and Jarvis, this issue). In particular, Rilling and colleagues found that while the macaque homolog of the arcuate fasciculus makes only specific and limited connections between the superior temporal gyrus and regions of the lateral prefrontal cortex, the human arcuate makes rich and extensive connections to the middle and inferior temporal gyri as well. This expansion of the connectivity from prefrontal regions to essentially the entire temporal lobe may be linked to a relative expansion in humans of both frontal (Deacon, 1997; Schoenemann et al., 2005) and temporal regions (Rilling and Seligman, 2002). This expanded connectivity was also found, but to a limited degree, in one of the four chimpanzee brains scanned.

The expansion and elaboration of the arcuate fasciculus in humans has a more interesting implication in the context of the current topic of exaptation, because many have noted that this pathway is involved not only in speech imitation but also in some aspects of syntax processing. Although Broca himself viewed his eponymous area as primarily involved in speech output, it has become abundantly clear that premotor areas such as BA 44 are also involved in auditory comprehension, and syntax processing in particular (e.g., Caramazza and Zurif, 1976; Grossman, 1980; Friederici et al., 2000). Because this issue is more thoroughly reviewed elsewhere in this issue, I will not go into the details here.

My main point is to suggest that these two functions – auditory/motor exchange and syntax production and comprehension – are not unrelated. Specifically, I hypothesize that reciprocal auditory/motor connections that initially evolved in the context of vocal imitation (an output function) could later be exapted for use in syntax comprehension (an input function). The logic underlying this suggestion is similar to that underlying the motor theory of speech production (Liberman et al., 1967; Liberman and Mattingly, 1985; Galantucci et al., 2006; Feenders et al., 2008; Pulvermüller and Fadiga, 2010) or more recent ideas about the role of the mirror neuron system in linking motor action and visual perception in the support of social cognition or empathy (di Pellegrino et al., 1992; Arbib, 2005; Iacoboni et al., 2005; Keysers and Gazzola, 2006). The prefrontal component of this system is, at birth, concerned basically with motor control. It develops, through babbling and later practice with speech production, a quite complex repertoire of automatized vocal actions. This initially occurs with auditory guidance, but has little significance for auditory comprehension itself. However, once such a learned cortical repertoire exists, it could provide a useful, highly articulated basis for auditory phonetic *comprehension* (knowing what a sound is based on the actions that would be needed to produce it) and, at a higher level, syntactic, or semantic (knowing what a sentence means based on the structures that would be needed to generate that meaning).

As phrased above, this hypothesis could apply either to ontogeny (during individual brain development) or phylogeny (during the evolution of the required neural circuits in the species). However, one reason to doubt a purely ontogenetic interpretation of syntactic circuitry is provided by studies of signed language (Bellugi et al., 1990). Although the output modality of sign is obviously different from that of speech (involving mostly limb movements, and monitored via vision and somatosensory systems), the syntactic processing of sign nonetheless relies on a brain region identical to, or at least greatly overlapping, Broca's area in spoken language (Horwitz et al., 2003). This suggests that the appropriate connections develop reliably in all humans, regardless of linguistic modality, and thus suggests an influence of an evolved pattern of connectivity in addition to the role of cortical plasticity during individual development.

Exaptive hypothesis 2 thus suggests that the intracortical connections of the human arcuate fasciculus initially evolved for the specific purpose of vocal imitation. These connections, once in place, were then again exapted for use in the more complex task of syntactic comprehension, and particularly between premotor control regions and posterior regions involved in semantic interpretation. This exaptation would have constituted a second evolutionary event, occurring afterward and perhaps for different selective reasons from the first.

EXAPTIVE HYPOTHESIS 3: THE EVOLUTION OF BROCA'S AREA (BA 44 AND 45)

The third exaptive hypothesis considered here is closely related to the previous one, and concerns the evolution of specific cortical regions involved in syntax. While the previous hypothesis is hodological, concerned with connections, this one is cytoarchitectonic, and concerned with the computational specializations of specific cortical regions. This hypothesis builds largely upon the work of Angela Friederici and her colleagues (Friederici et al., 2006; Anwander et al., 2007; Bahlmann et al., 2008). It is widely recognized that the left inferior frontal gyrus (LIFG), the traditional site of "Broca's area," is subdivided into multiple regions distinguishable on the basis of cytoarchitecture, receptor distributions, and connectivity (Amunts et al., 1999; Zilles, 2004). Four core regions can be distinguished, the first three designated by their BA:

1. The inferior portion BA 6 – premotor and supplementary cortex;
2. BA 44 – "pars opercularis" – just anterior to BA6;
3. BA 45 – "pars triangularis" – anterior to BA 44;
4. "Deep frontal operculum" – inferior to BA 44, and curling under the temporal lobe.

BA 44 and 45 are the core of Broca's area, as classically defined. In an important study, Horwitz et al. (2003) investigated brain activation during signing and speech by fluent bilinguals, and then correlated the brain activation patterns with cytoarchitectonic maps. They found that BA 44 was activated not only during speech or sign production, but also during complex, volitional movements of the limbs or vocal apparatus with no linguistic content. In contrast, BA 45 was activated only during the production of either spoken or signed language (complex narratives, so including phonological, syntactic, and semantic components). In both cases, activations were significantly above baseline only in the left hemisphere. These data suggest that Broca's region is parcellated into regions that differ in the degree to which they are tied to motor control.

The idea that Broca's area is made up of computationally distinct regions receives strong support from neuroanatomical research on both humans and monkeys (Rempel-Clower and Barbas, 2000; Anwander et al., 2007; Gerbella et al., 2010). Cytoarchitectonically, each of the four regions above is distinct in terms of the granularity of layer 4 of neocortex. BA 45 has a well-developed granular layer 4 (it is "granular"), while layer 4 is present but poorly developed in BA 44 (which is thus termed "dysgranular"). The deep frontal operculum and BA 6 are "agranular": they lack any distinct layer 4 (a trait typical of motor regions).

Connectivity also varies among these regions. Tracer studies in macaques, which remain the "gold standard" for analysis of connectivity, clearly show that differences in granularity reflect fundamental differences in connectivity (Rempel-Clower and Barbas, 2000), and such studies also show that BA 45 in macaques clearly divides into two distinct components. BA 45B is heavily connected to eye movement circuitry such as the frontal and supplementary eye fields. While BA 45A also makes extensive frontal connections, it is unique in having strong connections to superior temporal and auditory areas (Gerbella et al., 2010) and seems more directly homologous to human BA 45. Human tractography, using DTI and related methods, also reveals a clear parcellation of Broca's area (Anwander et al., 2007), with BA 44, BA 45, and the deep frontal operculum having distinct patterns of connectivity.

There are two distinct exaptive hypotheses about the evolutionary origins of language-specific functions in Broca's region, and particularly BA 45. The first, which I will dub the "premotor origins hypothesis," continues a long line of suggestions that the underlying computations of human language syntax are related in some way to motor control and motor planning. Many theorists have suggested that the hierarchical nature of linguistic structures is related in some way to the hierarchical nature of motor planning (e.g., Lashley, 1951; Lieberman, 1984; Kimura, 1993; Allott, 1994; MacNeilage, 1998). This suggestion leads to a more concrete suggestion in the current context: that the premotor functions of BA 6 and the deep frontal operculum (with their agranular layering pattern) provided the precursor of computations more directly characteristic of language through a gradual "granularization" of gray matter, and strengthening of pre-existing connections to other regions of cortex (e.g., via the arcuate fasciculus to parietal and temporal regions). Starting with BA 44, and progressing to BA 45, this would lead to a broader and more abstract computational role for Broca's area, culminating in the amodal, granular BA 45.

A second (and not mutually exclusive) hypothesis has not, to my knowledge, been previously suggested. Apes and Old World primates like chimpanzees and macaques have binocular vision and trichromacy, reflecting an increased importance of the visual modality relative to sound and olfaction. This is reflected specifically in a heightened awareness of the gaze of others, and volitional control of one's own eyes, both of which play important roles in Old World primate social behavior and understanding (Scaife and Bruner, 1975; Emery et al., 1997; Emery, 2000). For review of the importance of gaze and visual cueing in the evolution of cognition and language see (Fitch et al., 2010). While the movement of the eyes is clearly a motor function, its control requires strong intracortical communication from visual and multi-modal areas. When such visual dependency is combined with intense social pressures, we might expect the computation of eye movements to have a more abstract and generalized component than limb or hand movements. Since, in the macaque, one portion of BA 45 is closely linked to eye movements, while the other makes long-distance cortical connections, I suggest that the abstract and amodal computations involved in language (whether spoken or signed) had a pre-adaptive foundation in the social and visually guided aspects of gaze that are, by hypothesis, subserved by BA 45B.

These two hypotheses may in fact be complementary, in the sense that the special role of BA 44 in language production and processing may represent a kind of fusion of the two flanking regions, specifically the hierarchical premotor functions of BA 6, and the multi-modal, integrative, and social functions of BA 45. The result would be a more abstract computational process than hierarchical motor planning: an operator that can combine or unify pre-existing conceptual units (motor actions, vocalizations, or visual objects) to freely create a discrete infinity of modality-independent cognitive structures. Such an operator shares the key functionality required by modern theories of language, such as

Merge in minimalist approaches (Chomsky, 1995; Hendrick, 2003; Berwick and Chomsky, 2011) or "unify" in tree-adjoining, construction grammar or other unification grammars (cf. Hagoort, 2005b). Whether during comprehension or in production, such an operator must quickly retrieve items from memory (e.g., retrieve the phonological form of words from the lexicon), combine them in a context–relevant fashion (e.g., using background information and current context) into flexible, temporary, goal-relevant structures that can be parsed semantically (in comprehension) or produced motorically via some serialization process (during production). As emphasized by Hagoort, such an overarching computation is consistent with both the neuroanatomy of Broca's area, as discussed above, and a wide variety of brain imaging results focused on language comprehension (Hagoort, 2005a,b).

Could hierarchy-building circuitry in BA 6 and the deep frontal operculum, evolved in the context of motor planning and dedicated to motor control, be exapted to produce a general purpose, amodal, two-way circuit that can perform the computational equivalent of Merge or Unify? While this hypothesis clearly remains speculative at present, it has several points to recommend it:

1. We know from the brain imaging studies cited above that amodal, language circuitry exists in Broca's region and is involved in language production, regardless of output modality;
2. The modality-general portion of this circuitry appears to center on BA 45;
3. We know from both lesion and brain imaging studies that Broca's area more generally (BA 44 included) is involved not just in production by also in perception;
4. Studies of macaque mirror neurons and the human mirror system provide a plausible foundation for the two-way nature of this system, building on a pre-existing mirror system for interpreting the actions of others;
5. We know that Broca's area, broadly construed, can be activated in non-linguistic tasks, consistent with a generalized, amodal role of the LIFG in selection of appropriate action or items, and suppression of irrelevant alternatives.

Not only is this exaptive hypothesis consistent with the data above, but it makes several specific testable predictions about the structure, connectivity, and function of BA 45:

1. BA 45 in humans should be more amodal than other components of the LIFG, and in particular its white-matter connections should be longer, and fan out more widely, than those of other regions;
2. Anatomically, the cytoarchitecture of BA 45 should be less motor-driven, more perceptually embedded, and thus more suited to amodal cognition than BA 6 or BA 44 (as already suggested by its granular layer 4);
3. In monkeys, cells in BA 45A should fire in a much wider variety of situations than BA 44 or BA 6 (including in particular social cognition tasks).

I conclude that exaptive hypotheses about the origins of syntax, far from being "utterly implausible," are consistent with a wide range of neuroanatomical and functional data. I see such hypotheses, when built upon a solid foundation of comparative neuroscientific data, as presenting numerous testable predictions and avenues for profitable empirical investigation. Of course, "plausible" plus "testable" does not equal "true." But I hope to have convinced the open-minded reader that there are good reasons for considering exaptive hypotheses in research on the evolution of cognition, and that the mistrust or outright hostility such hypotheses have evoked from some quarters is unjustified.

TESTING EXAPTIVE HYPOTHESES: THE ROLE OF COMPARATIVE RESEARCH

I will end with a brief attempt to clarify the role of comparative research in testing evolutionary hypotheses in general, and exaptive hypotheses in particular. In brief, we can use comparison of homologous traits in closely related species to derive inferences about ancestral states. This is an important first step in testing any exaptive hypothesis: we must first understand the ancestral state that served as a putative "pre-adaptation." When considering the evolution of language, and neural components in particular, the comparison set here will typically be non-human primates, or sometimes mammals more generally.

The second set of comparisons involve "analogies" – similar traits that have evolved independently, via convergent evolution (such traits are one member of the broader class termed "homoplasies" by comparative biologists). Convergent evolution plays a central role in testing evolutionary hypotheses, because each example of a convergently evolved trait represents an independent evolutionary event, and thus an independent data point for statistical testing (Harvey and Pagel, 1991; Pagel, 1992). In contrast, a trait that is homologous among a group of species has, by definition, evolved only once in that clade, and even if there are hundreds of species sharing the trait, they constitute only a single independent data point.

The clade used most frequently in comparative tests regarding language evolution are birds, Class Aves, which have evolved numerous traits convergently with some mammals (e.g., homeothermy or "warm bloodedness," or large brains) and specifically with humans (e.g., biparental care, vocal imitation, or bipedalism).

HOMOLOGY: INFERENCES ABOUT ANCESTRAL STATE

A homologous trait is one that is shared among a group of related organisms by virtue of its presence in their common ancestor. When attempting to build a phylogeny, systematists distinguish between two classes of homology. "Synapomorphies" are traits that are shared in a particular clade, but lacking in other close relatives, while "symplesiomorphies" are more broadly shared (due to their present in a more ancient shared ancestor). Feathers are a synapomorphy among living bird species (all birds have them, and all living non-birds lack them), while the possession of a heart or a mouth are symplesiomorphies (all birds have them, but so do all other vertebrates). However, in the current context we are concerned with rebuilding ancestral states, however far back they might go, and so we will discuss homology in general.

Homology is a relative concept: it depends on what trait is being examined, and what particular clade is being discussed. Thus, for example, the wings of birds and bats are homologous *as forelimbs* (because they both derive from the forelimb of the shared tetrapod ancestor of birds and mammals) but are convergently evolved analogs *as wings*. Furthermore, correct determination of homology depends upon the level of mechanistic detail being discussed. The complex, image-forming eyes found in insects, octopus, or vertebrates evolved convergently (Allman, 1999) but their location is nonetheless controlled by a homologous transcription factor Pax-6 (Quiring et al., 1994; Gehring and Ikeo, 1999; van Heyningen and Williamson, 2002). Such a situation has been termed *deep homology* and appears to play a surprisingly important role in human evolution (cf. Carroll, 2008; Fitch, 2009; Shubin et al., 2009).

Careful examination of homologous traits in multiple species allows us to reconstruct traits that were present in the common ancestor of those species. For example, the corpus callosum is found in all placental mammals (eutherians), but is absent in marsupials and monotremes. This allows us to conclude that the corpus callosum was not present in the common ancestor of all mammals, but arose in the LCA of living eutherians. In contrast, the anterior commissure is found in marsupials, and more widely among vertebrates including birds, suggesting that it evolved rather early in tetrapods. Such inferences about ancestral states depend on solid comparative neuroanatomy in living organisms, and no fossil evidence is required to roughly date such evolutionary events.

A broad comparative analysis is also important to determine the *directionality* of any evolutionary changes in different lineages. It is important to note that traits can be lost as well as gained in evolution. There is no way to know, based on a simple comparison of two species that differ in some trait, what the directionality of change might have been, and this requires outgroup comparisons with other related species. For example, one might think that the sexual swellings surrounding the vaginal area in female chimpanzees are a primitive feature of primates, given that such swellings are also seen in macaques and baboons. This would lead to the conclusion that humans have lost such swellings in our recent evolutionary history. An outgroup comparison, looking at gorillas, orangutans, and gibbons (the other members of our own ape clade) refutes that intuitive inference: none of these apes shows sexual swellings. Thus in this case *humans* retain the primitive state, and it is chimpanzees that have evolved a novelty (convergently with moneys and baboons). This example illustrates both the need for rigorous outgroup comparisons when analyzing homology, and also illustrates that the LCA of humans and chimpanzees *was not a chimpanzee*. In this case, and presumably many others, chimpanzees have evolved, just as we have, in the last 6 million years.

CONVERGENCE: TESTING HYPOTHESES ABOUT MECHANISM AND ADAPTIVE VALUE

Homologies play a crucial role in constructing phylogenies and rebuilding extinct ancestors. For such problems, convergent evolution is a nuisance and a distraction. But for a different set of more abstract evolutionary problems, convergent evolution provides one of the most powerful tools at our disposal, and thus the second main arm of the comparative method. These are problems involving statistically valid tests of evolutionary hypotheses, whether of mechanism or evolutionary function.

In some cases convergent evolution of a similar trait can lead us to new hypotheses about function. Consider the example of the descended human larynx, which was until recently believed to be uniquely human, and therefore assumed to be directly tied to human speech abilities. The recent discoveries of a permanently descended larynx in several deer species, big cats such as lions and tigers, koalas, Mongolian gazelles, and goitered gazelles clearly demonstrates that a descended larynx can serve some purpose other than speech (Fitch, 2010). In this case, analyses strongly suggest that the purpose is an exaggeration of the size conveyed by vocalizations, so that animals with a descended larynx seem larger to listeners. Because this explanation also applies to humans, it offers a clear alternative to the standard idea that the descended larynx is an exclusively speech-related trait. This is particularly true since the human larynx descends a second time, at puberty, but only in males. This sexually dimorphic descent seems unlikely to be tied to speech abilities, since teenage males do not undergo any improvement in these: instead it is simply part of a deeper, richer, and more impressive voice timbre, quite similar to the red deer's roar.

In other cases convergent evolution of mechanism can corroborate mechanistic hypotheses. A nice example is given by the evolution of direct motor connections in birds. As discussed above in hypothesis one, a plausible reason chimpanzees and other primates cannot talk, and humans can, is because humans possess direct cortico-motor connections for laryngeal control, and other primates do not. This is a correlational argument: humans have speech and have direct connections, and other primates do not. Therefore, direct connections are involved in speech motor control. While plausible, such arguments are never particularly convincing. However, in this case, data from birds strongly support the argument, because direct connections from the avian equivalent of motor cortex (area RA) indeed have direct monosynaptic connections onto the motor neurons controlling the muscles of the syrinx (the birds primary organ of voice). These connections are lacking in bird species incapable of vocal learning (Wild, 1993a,b). This case, where an analogous mechanism is present in species which have convergently evolved the analogous ability, lends great credence to the Kuypers/Jürgens hypothesis. What is more, there are many other relevant species in which to test this prediction, including all the mammals who have convergently evolved vocal learning.

Many other examples of the power of convergent evolution could be drawn from the genetic level, already explored in previous work. For example, the role of FOXP2 in vocal control and vocal motor learning has been elegantly explored in birds by Constance Scharff and colleagues (Haesler et al., 2004, 2007; Scharff and Haesler, 2005), and provides the first example of deep homology yet known in the domain of vocal motor control (cf. Fitch, 2009; Scharff and Petri, 2011). Research on axon guidance has examined the development of a set of connections in the avian brain that may be the analogy of the arcuate fasciculus, and suggests that a set of broadly shared axonal guidance molecules, the cadherins, may play a key role in this (Matsunaga and Okanoya, 2008, 2009). An intriguing possibility is that these same molecules, or close relatives, might also play a role in the development of intracortical connections that underlie syntax or spoken language more broadly.

In summary then, both analogy and homology will always have a role to play in the comparative method, and the importance of convergent evolution for testing evolutionary hypotheses means that research on the avian brain will continue to be extremely important in analyses of traits that make humans unique among primates.

CONCLUSION

In this paper I have shown that models of neural/cognitive evolution that invoke exaptation – changes in function – are fully consistent with Darwin's ideas about evolution by natural selection, as well as with modern evolutionary biology and contemporary neuroscience. To illustrate this point I have assembled three exaptive hypotheses, focused on different mechanisms of human spoken language, to try to illustrate this approach from a "nuts and bolts" perspective, and to make some specific predictions that could be tested. Although I find each of these hypotheses plausible, I have no personal investment in them being true in any ultimate sense. But I do believe that they have a virtue that all good scientific hypotheses share: they are testable. Specific, testable hypotheses remain all too rare in research on the evolution of language. Furthermore, such hypotheses present numerous avenues for testing and elaboration using the comparative method, both for pinpointing ancestral states via homology, and for testing hypotheses using convergence.

Since testing plausible hypotheses is the surest route to progress I know of in science, I would be very pleased to see any of the three hypotheses above tested, even if refuted in the process. If that occurs, a new and better hypothesis will surely rise to take its place.

REFERENCES

Aitken, P. G. (1981). Cortical control of conditioned and spontaneous vocal behavior in rhesus monkeys. *Brain Lang.* 13, 171–184.

Allman, J. M. (1999). *Evolving Brains.* (New York: Scientifc American Library).

Allott, R. M. (1994). "Motor theory of language origin: the diversity of languages," in *Studies in Language Origins*, eds J. Wind, A. Jonker, R. M. Allott, and L. Rolfe (Amsterdam: John Benjamins), 125–160.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B. M., and Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412, 319–341.

Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., and Knoŝche, T. R. (2007). Connectivity-based parcellation of Broca's area. *Cereb. Cortex* 17, 816–825.

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* 28, 105–167.

Bahlmann, J., Schubotz, R. I., and Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *Neuroimage* 42, 525–534.

Bellugi, U., Poizner, H., and Klima, E. S. (1990). "Mapping brain funcion for language: evidence for sign language," in *Signal and Sense*, eds G. M. Edelman, W. E. Gall, and W. M. Cowan (New York, NY: John Wiley & Sons, Inc.), 521–543.

Berwick, R. C. (1997). Syntax facit saltum: computation and the genotype and phenotype of language. *J. Neurolinguistics* 10, 231–249.

Berwick, R. C., and Chomsky, N. (2011). "The biolinguistic program: the current state of its development," in *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, eds A. M. Di Sciullo and C. Boeckx (Oxford: Oxford University Press), 19–41.

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

Bickerton, D. (2000). "How protolanguage became language," in *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, eds C. Knight, M. Studdert-Kennedy, and J. R. Hurford (Cambridge: Cambridge University Press), 264–284.

Bickerton, D. (2007). Language evolution: a brief guide for linguists. *Lingua* 117, 510–526.

Bickerton, D. (2010). *Adam's Tongue*, New York: Hill and Wang.

Brown, S. (2000). "The "Musilanguage" model of music evolution," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 271–300.

Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A. L., and Wakefield, J. C. (1998). Adaptations, exaptations, and spandrels. *Am. Psychol.* 53, 533–548.

Byrne, R. (1997). Machiavellian intelligence. *Evol. Anthropol.* 5, 172–180.

Caramazza, A., and Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain Lang.* 3, 572–582.

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

Catani, M., and Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *Cortex* 44, 953–961.

Catania, K. C. (1999). A nose that looks like a hand and acts like an eye: the unusual mechanosensory system of the star-nosed mole. *J. Comp. Physiol. A* 185, 367–372.

Chomsky, N. (1968). *Language and Mind.* New York: Harcourt, Brace & World.

Chomsky, N. (1976). On the nature of language. *Ann. N. Y. Acad. Sci.* 280, 46–57.

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

Chomsky, N. (2010). "Some simple evo devo theses: how true might they be for language?" in *The Evolution of Human Language: Biolinguistic Perspectives*, eds R. Larson, V. Deprez, and H. Yamakido (Cambridge: Cambridge University Press), 45–62.

Clark, B. (2011). Scavening, the stag hun, and the evolution of language. *J. Linguist.* 47, 447–480.

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

Deacon, T. W. (1992). "The neural circuitry underlying primate calls and human language," in *Language Origins: A Multidisciplinary Approach*, eds J. Wind, B. A. Chiarelli, B. Bichakjian, and A. Nocentini (Dordrecht: Kluwer Academic), 301–323.

Deacon, T. W. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain.* New York: Norton.

Dennett, D. C. (1995). *Darwin's Dangerous Idea.* New York: Simon & Schuster.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604.

Emery, N. J., Lorincz, E. N., Perrett, D. I., Oram, M. W., and Baker, C. I. (1997). Gaze following and joint attention in rhesus monkeys (*Macaca mulatta*). *J. Comp. Psychol.* 111, 286–293.

Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., Wada, K., Mouritsen, H., and Jarvis, E. D. (2008). Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS ONE* 3, e1768. doi:10.1371/journal.pone.0001768

Fitch, W. T. (2002). "Comparative vocal production and the evolution of speech: reinterpreting the descent of the larynx," in *The Transition to Language*, ed. A. Wray (Oxford: Oxford University Press), 21–45.

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215.

Fitch, W. T. (2009). "The biology & evolution of language: "deep homology" and the evolution of innovation," in *The Cognitive Neurosciences IV*, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 873–883.

Fitch, W. T. (2010). *The Evolution of Language.* Cambridge: Cambridge University Press.

Fitch, W. T., Huber, L., and Bugnyar, T. (2010). Social cognition and the evolution of language: constructing cognitive phylogenies. *Neuron* 65, 795–814.

Fitch, W. T., and Reby, D. (2001). The descended larynx is not uniquely human. *Proc. Biol. Sci.* 268, 1669–1675.

Foerster, O. (1936). "Motorische Felder und Bahnen," in *Handbuch der Neurologie*, eds Band IV, O. Bumke, and O. Foerster (Berlin: Springer), 1–448.

Frey, R., and Riede, T. (2003). Sexual dimorphism of the larynx of the Mongolian Gazelle (*Procapra gutturosa* Pallas, 1777) (Mammalia, Artiodactyla, Bovidae). *Zool. Anz.* 242, 33–62.

Friederici, A., Meyer, M., and von Cramon, D. Y. (2000). Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. *Brain Lang.* 74, 289–300.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., and Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2458–2463.

Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377.

Gehring, W. J., and Ikeo, K. (1999). Pax 6: mastering eye morphogenesis and eye evolution. *Trends Genet.* 15, 371–377.

Gerbella, M., Belmalih, A., Borra, E., Rozzi, S., and Luppino, G. (2010). Cortical connections of the macaque caudal ventrolateral prefrontal areas 45A and 45B. *Cereb. Cortex* 20, 141–168.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. *Brain* 88, 585–644.

Geschwind, N. (1970). The organization of language and the brain. *Science* 170, 940–944.

Gould, S. J. (1991). Exaptation: a crucial tool for evolutionary psychology. *J. Soc. Issues* 47, 43–65.

Gould, S. J., and Vrba, E. S. (1982). Exaptation – a missing term in the science of form. *Paleobiology* 8, 4–15.

Grossman, M. (1980). A central processor for hierarchically-structured material: evidence from Broca's aphasia. *Neuropsychologia* 18, 299–308.

Groswasser, Z., Korn, C., Groswasser-Reider, J., and Solzi, P. (1988). Mutism associated with buccofacial apraxia and bihemispheric lesions. *Brain Lang.* 34, 157–168.

Haesler, S., Rochefort, C., Geogi, B., Licznerski, P., Osten, P., and Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. *PLoS Biol.* 5, e321. doi:10.1371/journal.pbio.0050321

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D., and Scharff, C. (2004). FoxP2 expression in avian vocal learners and non-learners. *J. Neurosci.* 24, 3164–3175.

Hagoort, P. (2005a). "Broca's complex as the unification space for language," in *Twenty-First Century Psycholinguistics: Four Cornerstones*, ed. A. Cutler (London: Lawrence Erlbaum), 157–172.

Hagoort, P. (2005b). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423.

Harvey, P. H., and Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.

Hayes, C. (1951). *The Ape in Our House.* New York: Harper.

Hendrick, R. (ed.). (2003). *Minimalist Syntax.* Oxford: Blackwell.

Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Curr. Anthropol.* 14, 5–24.

Holstege, G., and Ehling, T. (1996). "Two motor systems involved in the production of speech," in *Vocal Fold Physiology: Controlling Complexity and Chaos*, eds P. J. Davis and N. H. Fletcher (San Diego: Singular Press), 153–196.

Horwitz, B., Amunts, K., Bhattacharyya, R., Patkin, D., Jeffries, K., Zilles, K., and Braun, A. R. (2003). Activation of Broca's area during the production of spoken and signed language: a combined cytoarchitectonic mapping and PET analysis. *Neuropsychologia* 41, 1868–1876.

Humphrey, N. K. (1976). "The social function of intellect," in *Growing Points in Ethology*, eds P. P. G. Bateson and R. A. Hinde (Cambridge: Cambridge University Press), 303–317.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3, e79. doi:10.1371/journal.pbio.0030079

Jackendoff, R., and Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, & Chomsky). *Cognition* 97, 211–225.

Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166.

Jacob, F. (1982). *The Possible and the Actual.* New York: Pantheon.

Jolly, A. (1966). Lemur social behavior and primate intelligence. *Science* 153, 501–506.

Jürgens, U. (1994). The role of the periaqueductal grey in vocal behaviour. *Behav. Brain Res.* 62, 107–117.

Jürgens, U., Kirzinger, A., and von Cramon, D. Y. (1982). The effects of deep-reaching lesions in the cortical face area on phonation: a combined case report and experimental monkey study. *Cortex* 18, 125–139.

Keysers, C., and Gazzola, V. (2006). Towards a unifying neural theory of social cognition. *Prog. Brain Res.* 156, 379–401.

Kimura, D. (1993). *Neuromotor Mechanisms in Human Communication.* Oxford: Oxford University Press.

Kingsolver, J. G., and Koehl, M. A. R. (1985). Aerodynamics, thermoregulation, and the evolution of insect wings: differential scaling and evolutionary change. *Evolution* 39, 488–504.

Kuypers, H. G. J. M. (1958a). Corticobulbar connections to the pons and lower brainstem in man: an anatomical study. *Brain* 81, 364–388.

Kuypers, H. G. J. M. (1958b). Some projections from the pericentral cortex to the pons and lower brain stem in monkey and chimpanzee. *J. Comp. Neurol.* 110, 221–255.

Kuypers, H. G. J. M. (1973). "The anatomical organization of the descending pathways and their contributions to motor control especially in primates," in *New Developments in EMG and Clinical Neurophysiology*, ed. J. E. Desmedt (Basel: Karger), 38–68.

Lashley, K. (1951). "The problem of serial order in behavior," in *Cerebral Mechanisms in Behavior; The Hixon Symposium*, ed. L. A. Jeffress (New York: Wiley), 112–146.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461.

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36.

Lieberman, P. (1968). Primate vocalization and human linguistic ability. *J. Acoust. Soc. Am.* 44, 1574–1584.

Lieberman, P. (1975). *On the Origins of Language.* New York: Macmillan.

Lieberman, P. (1984). *The Biology and Evolution of Language.* Cambridge, MA: Harvard University Press.

Lieberman, P. (2000). *Human Language and Our Reptilian Brain: the Subcortical Bases of Speech, Syntax and Thoug*ht. Cambridge, MA: Harvard University Press.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21, 499–546.

Marshall, A. J., Wrangham, R. W., and Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees? *Anim. Behav.* 58, 825–830.

Matsunaga, E., and Okanoya, K. (2008). Expression analysis of cadherins in the songbird brain: relationship to vocal system development. *J. Comp. Neurol.* 508, 329–342.

Matsunaga, E., and Okanoya, K. (2009). Evolution and diversity in avian vocal system: an Evo-Devo model from the morphological and behavioral perspectives. *Dev. Growth Differ.* 51, 355–367.

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*. London: Weidenfeld & Nicolson.

Myers, R. E. (1976). Comparative neurology of vocalization and speech: proof of a dichotomy. *Ann. N. Y. Acad. Sci.* 280, 745–757.

Newmeyer, F. J. (1998). "On the supposed 'counterfunctionality' of universal grammar: some evolutionary implications," in *Approaches to the Evolution of Language*, eds J. R. Hurford, M. Studdert-Kennedy, and C. Knight (New York: Cambridge University Press), 305–319.

Nimchinsky, E. A., Gilissen, E., Allman, J. M., Perl, D. P., Erwin, J. M., and Hof, P. R. (1999). A neuronal morphologic type unique to humans and great apes. *Proc. Natl. Acad. Sci. U.S.A.* 96, 5268–5273.

Pagel, M. D. (1992). A method for the analysis of comparative data. *J. Theor. Biol.* 156, 434–442.

Piattelli-Palmarini, M. (1994). Ever since language and learning: afterthoughts on the Piaget-Chomsky debate. *Cognition* 50, 315–346.

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

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

Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360.

Quiring, R., Walldorf, U., Kloter, U., and Gehring, W. J. (1994). Homology of the eyeless gene of *Drosophila* to the small eye gene in mice and aniridia in humans. *Science* 265, 785–789.

Reeve, H. K., and Sherman, P. (1993). Adaptation and the goals of evolutionary research. *Q. Rev. Biol.* 68, 1–32.

Rempel-Clower, N. L., and Barbas, H. (2000). The laminar pattern of connections between prefrontal and anterior temporal cortices in the rhesus monkey is related to cortical structure and function. *Cereb. Cortex* 10, 851–865.

Reynolds Losin, E. A., Russell, J. L., Freeman, H., Meguerditchian, A., and Hopkins, W. D. (2008). Left hemisphere specialization for oro-facial movements of learned vocal signals by captive chimpanzees. *PLoS ONE* 3, e2529. doi:10.1371/journal.pone.0002529

Richman, B. (1993). On the evolution of speech: singing as the middle term. *Curr. Anthropol.* 34, 721–722.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., and Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11 4, 426–428.

Rilling, J. K., and Seligman, R. A. (2002). A quantitative morphometric comparative analysis of the primate temporal lobe. *J. Hum. Evol.* 42, 505–533.

Roe, A. W., Pallas, S. L., Kwon, Y. H., and Sur, M. (1992). Visual projection routed to the auditory pathway in ferrets: receptive fields of visual neurons in primary auditory cortex. *J. Neurosci.* 12, 3651–3664.

Scaife, M., and Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature* 253, 265–266.

Scharff, C., and Haesler, S. (2005). An evolutionary perspective on FoxP2: strictly for the birds? *Curr. Opin. Neurobiol.* 15, 694–703.

Scharff, C., and Petri, J. (2011). Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philos. Trans. R. Soc. Lond. B* 366, 2124–2140.

Schoenemann, P. T., Sheehan, M. J., and Glotzer, L. D. (2005). Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nat. Neurosci.* 8, 242–253.

Shubin, N., Tabin, C., and Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature* 457, 818–823.

Striedter, G. F. (2004). *Principles of Brain Evolution.* Sunderland, MA: Sinauer.

Suga, N., Niwa, H., Taniguchi, I., and Margoliash, D. (1987). The personalized auditory cortex of the mustached bat: adaptation for echolocation. *J. Neurophysiol.* 58, 643–654.

Sutton, D., Larson, C., and Lindeman, R. C. (1974). Neocortical and limbic lesion effects on primate phonation. *Brain Res.* 71, 61–75.

van Heyningen, V., and Williamson, K. A. (2002). PAX6 in sensory development. *Hum. Mol. Genet.* 11, 1161–1167.

Weissengruber, G. E., Forstenpointner, G., Peters, G., Kübber-Heiss, A., and Fitch, W. T. (2002). Hyoid apparatus and pharynx in the lion (*Panthera leo*), jaguar (*Panthera onca*), tiger (*Panthera tigris*), cheetah (*Acinonyx jubatus*), and domestic cat (*Felis silvestris f. catus*). *J. Anat.* 201, 195–209.

Wernicke, C. (1874). *Der Aphasische Symptomenkomplex.* Breslau: Cohn and Weigert.

Wild, J. M. (1993a). The avian nucleus retroambigualis: a nucleus for breathing, singing and calling. *Brain Res.* 606, 119–124.

Wild, J. M. (1993b). Descending projections of the songbird nucleus robustus archistriatalis. *J. Comp. Neurol.* 338, 225–241.

Yerkes, R. M., and Yerkes, A. W. (1929). *The Great Apes.* New Haven, CT: Yale University Press.

Zilles, K. (2004). "Architecture of the human cerebral cortex: regional and laminar organization," in *The Human Nervous System*, eds G. Paxinos and J. K. Mai (Amsterdam: Elsevier), 997–1055.

Conflict of Interest Statement: The author declares 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: 03 November 2011; accepted: 02 December 2011; published online: 23 December 2011.

Citation: Fitch WT (2011) The evolution of syntax: an exaptationist perspective. Front. Evol. Neurosci. ***3****:9. doi: 10.3389/fnevo.2011.00009*

frontiers in
EVOLUTIONARY NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 21 March 2012
doi: 10.3389/fnevo.2012.00004

The language–number interface in the brain: a complex parametric study of quantifiers and quantities

Stefan Heim[1,2,3,4]*, Katrin Amunts[1,2,4], Dan Drai[5], Simon B. Eickhoff[1,2,4,6], Sarah Hautvast[2] and Yosef Grodzinsky[2,7,8]

[1] Section Structural-functional Brain Mapping, Department of Psychiatry, Psychotherapy, and Psychosomatics, Medical School, RWTH Aachen University, Aachen, Germany
[2] Research Centre Jülich, Institute of Neuroscience and Medicine (INM-1 and INM-2), Jülich, Germany
[3] Section Neurological Cognition Research, Department of Neurology, Medical School, RWTH Aachen University, Aachen, Germany
[4] Jülich-Aachen Research Alliance – Translational Brain Medicine, Jülich and Aachen, Germany
[5] School of Communication, Inter Disciplinary Center, Herzylia, Israel
[6] Institut für Klinische Neurowissenschaften und Medizinische Psychologie, Heinrich-Heine Universität, Düsseldorf, Germany
[7] Department of Linguistics, McGill University, Montreal, QC, Canada
[8] Department of Neurology/Neurosurgery, McGill University, Montreal, QC, Canada

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

Reviewed by:
Angela Dorkas Friederici, Max Planck Institute for Human Cognitive and Brain Sciences, Germany
Paul Summers, University of Modena, Italy

****Correspondence:***
Stefan Heim, Department of Psychiatry, Psychotherapy, and Psychosomatics, Medical School, RWTH Aachen University, Pauwelsstraße 30, 52074 Aachen, Germany.
e-mail: sheim@ukaachen.de

The neural bases for numerosity and language are of perennial interest. In monkeys, neural separation of numerical Estimation and numerical Comparison has been demonstrated. As linguistic and numerical knowledge can only be compared in humans, we used a new fMRI paradigm in an attempt to dissociate Estimation from Comparison, and at the same time uncover the neural relation between numerosity and language. We used complex stimuli: images depicting a proportion between quantities of blue and yellow circles were coupled with sentences containing quantifiers that described them (e.g., "most/few of the circles are yellow"). Participants verified sentences against images. Both Estimation and Comparison recruited adjacent, partially overlapping bi-hemispheric fronto-parietal regions. Additional semantic analysis of positive vs. negative quantifiers involving the interpretation of quantity and numerosity specifically recruited left area 45. The anatomical proximity between numerosity regions and those involved in semantic analysis points to subtle links between the number system and language. Results fortify the homology of Estimation and Comparison between humans and monkeys.

Keywords: intraparietal sulcus, numerosity, semantics, fMRI, Broca's area, BA 45, estimation

INTRODUCTION

No species but ours can pride itself in the possession of both mathematical and linguistic abilities. Some primates possess numerical abilities, but only we humans can talk about what we count. Are these two abilities governed by the same principles and supported by the same neural structures? Philosophers from Plato to Frege have pondered the relation between mathematical and natural languages, as these seem to share properties, having alphabets and combinatorial rules, allowing for recursion, as well as ambiguous expressions. Chomsky (1988), for instance, has speculated that "the number faculty developed as a by-product of the language faculty" (p. 169; see Changeux and Connes, 1995, for a recent similar position). In contrast to these speculations, experimental evidence, mostly from double-dissociations observed in brain-damaged patients, has suggested that neurologically, the two capacities are distinct (Dehaene and Cohen, 1997; Cohen and Dehaene, 2000; Brannon, 2005; Gelman and Butterworth, 2005).

We report an fMRI experiment that addressed the language/math question from a new angle, and helped uncover the anatomical loci of linguistic and numerical operations. Our behavioral and imaging data seem to provide a fresh perspective on this perennial debate. Guided by current models of linguistic and mathematical capacity, we separated the neural underpinnings of complex language processes from those of numerosity-related ones. Within the latter, moreover, we were successful in identifying two distinct neural systems (one for quantity estimation, the other, for comparison). This three-way distinction between neurocognitive components emerged as subjects were evaluating linguistic statements about quantity against visual scenarios. Critically, it emanated from analyses of data from a single cross-modal parametric fMRI experiment that concomitantly probed numerical cognition and language.

The current view on numerical cognition and its brain basis is rather refined (Dehaene and Cohen, 1997; Butterworth, 1999; Cantlon and Brannon, 2007), drawing on complex computational models, and on empirical results from humans and monkeys. These models parse basic arithmetical operations into distinct steps – notably estimation (the encoding of numerical size), and comparison (the calculation of numerical distance), driven by Weber's Law, that states that the ability to perceive a minimal difference between the quantities of two stimuli is invariant across the range of possible stimulus sizes. The consequence of this law in the present context is, roughly, that the organism's ability to compare two quantities is a function of the proportion, rather than the distance, between them (Meck and Church, 1983; Dehaene and Changeux, 1993; Verguts and Fias, 2004). Experiments on numerosity typically present a sequence of images to participants. The first image depicts a number of objects (the *reference* numerosity *r*), and is followed by other images, of which the crucial one contains the *comparandum* numerosity *c*. In most instances, *r* is kept constant across trials, while *c* is a systematically varied parameter. Typically, this sequence is not accompanied by linguistic material.

The implementation of tasks that require the estimation of *r* and *c*, and then *r:c* comparison, has led to the discovery of neural correlates of these two operations in the macaque brain. There, neurons in lateral intraparietal cortex (LIP) in both hemispheres that respond to a relatively broad range of numerosities were distinguished from anteriorly adjacent (VIP) and frontal neurons that code for numerical distance, namely carry out comparisons (Nieder and Miller, 2003; Piazza et al., 2004; Roitman et al., 2007). Monkeys' numerical judgment behavior is, moreover, said to be asymmetrical, best described by Gaussian curves on a logarithmically compressed scale, whose variability is fixed across numerosities (Dehaene et al., 2003).

In humans, bilateral parietal and frontal regions modulated by numerical distance have also been identified, and shown to have similar properties: in a seminal study by Piazza et al. (2004), participants viewed a sequence of images that contained a fixed number of objects *r*, thereby adapting to *r*; the critical part then presented an image with the same objects, but whose numerosity *c* was parameterized. Using an fMRI adaptation paradigm (i.e., using the habituation of the fMRI signal as an indicator of repeated involvement of a region in the same kind of cognitive process), they found regions in right angular gyrus, right intraparietal sulcus (IPS), and right superior parietal lobe (SPL) that follow the relation between *r* and *c*, indicating involvement in numerical comparison: These regions present an adaptation response pattern to a numerical distance parameter (manipulated in a comparison task) that resembles monkeys' behavioral and neural patterns.

For language to be directly related to numerical cognition, aspects of language processing that pertain to quantity would be expected to have a role in numerical cognition. Past studies have either investigated linguistic and arithmetical abilities separately Dehaene et al., 2003), or focused on the relation between number words and their denotation (Dehaene and Cohen, 1997; Cohen and Dehaene, 2001; Varley et al., 2005). Investigations into the relation between the processing of sentences about numerical relations and corresponding number-containing scenarios are few and far between (McMillan et al., 2005; Troiani et al., 2009), despite the fact that we as speakers do just this naturally and frequently, as we evaluate statements we hear (or speak) against their real-life contexts. As a result, how exactly the neural substrate for numerical calculation relates to language is not entirely clear. Our project therefore involved a rather natural task, one that required participants to relate linguistic expressions to scenarios whose analysis necessitated numerical Estimation and Comparison: the scenarios contained quantities of objects, whereas the linguistic materials were sentences with quantity denoting expressions, that is, natural language quantifiers like *many*, *few*, *most*, or *less-than-half*. Successful performance required proper parsing of the visual scenarios, as well as an analysis of sentence meaning. Our contrasts, then, involved quantifiers that differed in meaning from one another.

We probed circuits for linguistic and mathematical processing in healthy participants in an fMRI experiment whose design featured two parameters, one that served as a proxy for numerical estimation, and the other, for comparison. It also featured a linguistic contrast that was orthogonal to numerosity (Polarity, the difference between positive quantifiers like *more-than-half* and *less-than-half*), which we used to detect certain aspects of the analysis of linguistic meaning which are orthogonal to numerosity. This design enabled us to measure the BOLD response related to each of these operations separately on the same data set.

A parametric proportion paradigm (PPP henceforth) was introduced: participants were asked to verify auditory sentences about proportion against visual scenarios that contained two quantities of objects, and do so under time pressure. The relevant meaning representations were created through the inclusion of a proportional (or degree) quantifier in each sentence, which was either positive, e.g., *many of the circles are blue*, or negative, e.g., *few of the circles are blue*. Each sentence was immediately followed by a visual image with an array of 50 quasi-randomly positioned circles. While the number of circles was fixed at 50 (a numerosity chosen to avoid subitizing), their radii were varied, and they were divided into two contiguous clusters of blue and yellow circles. The truth or falsity of each sentence depended on the meaning of the sentence and its match to the blue/yellow proportion (**Figure 1A**). To verify the sentence above, for example, participants had to first estimate *c*, the number of circles in the target color mentioned in the sentence (here blue), and *r*, the number of circles in the other color (here yellow). Next, they had to compare *c* to *r*. They then had to judge the truth-value of the sentence against the representation of the scenario (semantic truth-value judgment, SVJT). A sentence was deemed true just in case if $c > r$. Alternatively, as $r + c = 50$ in every case, a comparison of *c* to $25 = 50/2$ would also suffice, as they sentence is true when $c > 25$. Either way, the successful completion of the task in a limited time window required both estimation and comparison. BOLD response as well as reaction time (RT) were measured.

The PPP design systematically varied the proportion between the colors across scenarios. Participants thus judged each sentence against eight different scenarios (created in Mathematica™), each with a different blue/yellow proportion (**Figure 1B**, Properties of the Images and How They Were Created in Supplementary Material). The manipulation of *c* (= *Target Color, TarCol*) modulated RT: a change in *c* it coupled to a change in *r* (because $c + r = 50$), and affects the numerical distance between *c* and *r*; this, in turn, affects the difficulty of the comparison component of the task – the closer *c* is to *r* (or to 25), the more difficult the task becomes, and RT is elevated relative to trials in which *c* is distant from *r* (or from 25). The PPP used sentences conveyed weak statements that are true in multiple scenarios, and so truth-value was kept constant across several values of *c*. Moreover, they were contrasted along the linguistic dimension of Polarity, as some contained negative quantifiers (*few*) whereas others were positive (*many*).

This design allows correlating the BOLD response separately with two parameters, and one contrast:

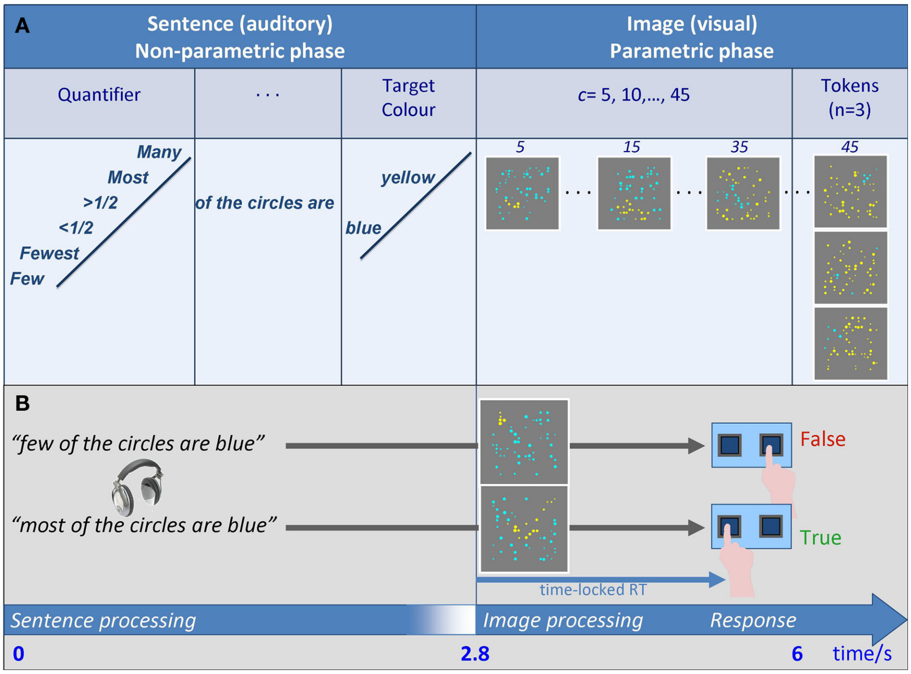


FIGURE 1 | Experimental design. (A) Overall design: Auditory sentence stimuli were of the form "*Quantifier* of the circles are *color*," where *Quantifier* was one of six proportional quantifiers, and *color* was blue or yellow. Each of the 12 resulting sentences was followed by a visual array of a fixed numerosity $T = 50$ blue and yellow circles, in which blue/yellow ratio was varied parametrically: the comparandum numerosity c of circles in the *TarCol* was a parameter, taking the values 5, 10, . . ., 45. As T was fixed, the numerosity of circles in the other color *r* varied accordingly (45, 40, . . ., 5 out of 50). Three different tokens were generated for each proportion, and combined with the sentences. **(B)** Time course of a trial: The auditory part was 2.8 s long (sentence duration was ≤2.8 s). Subsequently, an image was presented (1.2 s). Participants were asked to make a truth-value judgment by a button press. RTs were time-locked to the visual phase of the trial, where overall trial duration was 6 s. Speeded responses were mostly <2 s, which enabled a short rest period before the next trial began.

- The Estimation parameter *c* (i.e., Number of elements of *Target Color*): an increase in *c* forces the estimation of a larger numerosity. Changing *c* should incrementally tax the process of coding for numerical size, and activate brain regions that support it. Thus *c* can be used as a proxy to Estimation, and to identify regions involved in numerical estimation.
- The Comparison parameter RT: The closer *c* gets to *r*, the more difficult the comparison becomes, affecting RT (**Figure 2**). RT can therefore be used as an index of numerical comparison. Individual trial RTs were therefore used to identify regions involved in the comparison between *c* and *r* (or *c* and 25).
- The Semantic Polarity contrast: linguistic stimuli were divided along a clear non-numerical axis that related to their meaning – Polarity, characterized by whether or not the meaning representation of the quantifier contained a negation (cf. The Meaning of the Proportional Quantifiers Used in Supplementary Material for details). Half of the quantifiers were positive (*many, most, more-than-half*), and the other half negative (*few, fewest, less-than-half*). Contrasting these two types revealed brain regions involved in the linguistic analysis of quantifier meaning during the PPP task. Crucially, this contrast is orthogonal to both Estimation and Comparison, ensuring that numerosity circuits are excluded.

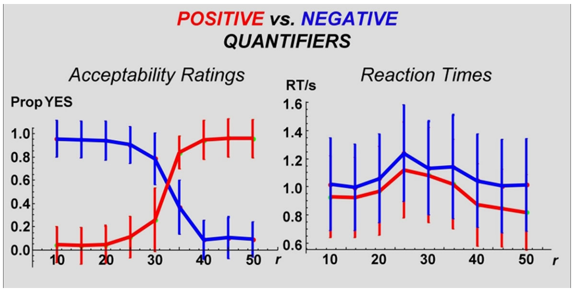


FIGURE 2 | Response accuracy and response time: Accuracy: Probabilities of "YES" responses by *TarCol* proportion per quantifier. NB: Quantifiers are clustered by Polarity (positive vs. negative). In both cases, a step-function is evident. Moreover, positive and negative quantifiers reveal a mirrored pattern – an image that makes a sentence with a positive quantifier true falsifies its negative counterpart. RT: Mean RTs by *TarCol* proportion per quantifier. NB: Quantifiers are clustered by Polarity. Negative quantifiers produce longer *RTs* (for statistical analysis see text). In all cases, a non-monotonic function is evident (see The Mapping Between the Current PPP and Standard Numerosity Experiments in Supplementary Material for mathematical details).

The experimental goal, then, was to obtain a three-way dissociation within the same fMRI data set. This goal was accomplished by identifying voxels responding to the one or the other of the uncorrelated parameters (c the monotonic estimation parameter, and RT, the non-monotonic comparison parameter), or to the semantic contrast.

The present design differs from previous numerosity experiments, and experiments that used linguistic stimuli to study numerical cognition (Cohen and Dehaene, 2001; McMillan et al., 2005; Troiani et al., 2009) by the following major aspects: first, we used two (c, RT) instead of a single parameter; second, a variable comparandum c is typically compared to a *fixed* reference numerosity r, and so $r+c$ is not constant. As we used scenarios that were suited to express proportion, r and c were related, such that $r+c=50$. Our paradigm and the standard one, however, are inter-translatable, and the relationship between them is formally well-defined (see The Mapping Between the Current PPP and Standard Numerosity Experiments in Supplementary Material for a mathematical translation between the two paradigms). Finally, the language materials included six different proportion or degree quantifiers – words like *many*, *most*, and *few* – that (1) denote relations between cardinalities of sets (as opposed to specific numbers), (2) that presuppose both estimation and comparison, and (3) that are weak, in that their truth-value is preserved across multiple scenarios. This last property makes them suitable to parametric designs, because it allows a sentence that contains such a quantifier to keep its truth-value constant across different numerosities (**Table 1**). We used several quantifiers, distinguished along the Polarity dimension (positive vs. negative), which enabled us to correlate the BOLD signal with a contrast that is purely linguistic, and does not bear on numerosity, even if both sides of this contrast come into contact with numerosity-related scenarios (see Materials and Methods , and The Meaning of the Proportional Quantifiers Used in Supplementary Material for further discussion).

MATERIALS AND METHODS

All procedures were approved by the ethics committee of the Medical School at RWTH Aachen University.

MATERIALS AND TASK

Participants performed a truth-value judgment task, indicating by pressing the left or right response button whether an auditory sentence with a quantified subject matched a subsequently presented visual array of blue and yellow circles. Each sentence was presented 24 times, with one picture at a time. The 24 repetitions result from systematically combining each sentence with three different pictures for each of the eight different proportions of blue/yellow circles (see below). Stimulus presentation was controlled by a computer placed in the control room using Presentation 11.0 software (Neurobehavioral Systems, Albany, CA, USA), and each participant received a different pseudo-randomization of sentence–picture pairings.

Table 1 | German quantifiers (with English translations) used in the study.

Quantifier	Polarity
Viele = many	Positive
Wenige = few	Negative
Die meisten = most	Positive
Die wenigsten = very few	Negative
Mehr als die Hälfte = more-than-half	Positive
Weniger als die Hälfte = less-than-half	Negative

Stimuli were constructed of auditory sentence–visual image pairs (cf. **Figure 1**). Images depicted a scenario with 50 circles, divided unequally between yellow and blue, with the constraint that each set of colored circles form a cluster. The six experimental conditions contained sentences that were built out of well-defined pieces, all containing a proportional quantifier that composed with a restrictor noun (*Kreise* = circles, in our case) to form a generalized quantifier. This partitive expression (e.g., *Viele der Kreise* = many of the circles) was the subject of a copular sentence with a color predicate adjective (*blau* = blue or *gelb* = yellow – the *Target Color*, resulting in sentences such as *Viele die Kreise sind gelb* = many of the circles are yellow (Milsark, 1977; Diesing, 1992; see The Mapping Between the Current PPP and Standard Numerosity Experiments in Supplementary Material).

The trial schema (**Figure 1**) involved the presentation of the sound file containing one of the stimulus sentences. Sound files had a duration of below 2.8 s. At a trial time of 2.8 s after the onset of the sound file, the visual scenario containing blue and yellow circles was presented. The subjects' responses were time-locked to the onset of the visual scenario. Each trial had an overall duration of 6 s, after which time the next trial started.

fMRI DATA ACQUISITION

The fMRI experiment was carried out on a 3-T Trio scanner (Siemens, Erlangen, Germany). A standard birdcage head coil was used with foam paddings to reduce head motion. The functional data were recorded from 40 axial slices using a gradient-echo EPI sequence with echo time (TE) = 30 ms, flip angle = 90°, and repetition time (TR) = 3 s. The field of view (FOV) was 256 mm, with an in-plane resolution of 3 mm × 3 mm. The slice thickness was 3 mm with an inter-slice gap of 1 mm. A time series containing a total of 684 images was recorded, amounting to a total functional scanning time of 34 min.

fMRI DATA ANALYSIS

Pre-processing

Data analysis was performed using SPM5 (Wellcome Department of Cognitive Neurology, UK) running on MATLAB 7 (The Mathworks Inc., Natick, MA, USA). Pre-processing involved the standard procedures of realignment to the mean image of the EPI time series, normalization of functional data to the MNI template using the unified segmentation procedure provided in SPM5, spatial smoothing with a Gaussian kernel of 8 mm FWHM, and highpass filtering at 1/128 Hz in order to correct for slow drifts in the BOLD signal.

Statistical analyses

Each of the different processing steps involved in quantification, i.e., Composition, Estimation, and Comparison, was addressed by its own event-related analysis (note that the differential analysis of the auditory phase vs. the visual phase in a trial is enabled by the temporal spacing of the onset times of more than 1 s: Boynton et al., 1996; Dale and Buckner, 1997). Moreover, subsequent analysis of Semantic Structure was performed. Data from individual first level analyses were then entered into random-effects group analyses at the second level using the *flexible factorial* option for repeated-measures designs in SPM5.

1. In a first analysis for Composition, the set of brain regions was assessed which was activated while listening to the auditory sentence that contained the quantifying expression and building up a semantic representation of the according scenario (data type: beta estimates of BOLD amplitude during the auditory presentation phase).
2. The next, parametric, analysis was run for Estimation, i.e., coding the circles in the *TarCol* for numerical size (data type: beta estimates of the (monotonically increasing) parametric modulator of the amount of circles of the *TarCol* during the visual presentation phase).
3. The third, again parametric, analysis was done for Comparison, i.e., coding for numerical distance between |c| and its complement color (data type: beta estimates of the (non-monotonic) RT regressor during the visual presentation phase) and comparing it to the mental representation created during Composition. To this end, we identified voxels which (1) parametrically responded to the RT regressor and (2) were contained in the Composition phase during which the mental representation was first generated. The resulting Comparison effect was thus calculated by a conjunction analysis of second level T-statistic maps obtained from Composition and the RT parameter. Since the effect for Composition was obtained in the auditory domain and that for the RT regressor in the visual domain, the conjunction analysis reveals such areas that are involved in a-modal processing of semantic representations not solely related to the auditory sentence or the visual scenario. This type of analysis is valid despite the different nature of the raw data, since it is calculated on the T-statistic maps which are all at the same (T) scale.

Composition. Composition (i.e., creation of a semantic representation) refers to the first, auditory phase of the trial when subjects listened to a sentence containing a quantifier expression. There were six conditions, i.e., one for each quantifier.

1. At the first level, an event-related general linear model (GLM) analysis was performed. The duration for each condition was set to 2.8 s, beginning with sentence-onset and covering the entire auditory phase. Each condition was convolved with the canonical hemodynamic response function (HRF) and its first temporal derivative. For subsequent ANOVA at the second level, the beta weights for the six Composition conditions were contrasted against the implicit resting baseline by calculating contrasts of the type `'1 0 0 0 0 0'`.
2. At the second (group) level, the random-effects repeated-measures 1 × 6 ANOVA was calculated in order to obtain an T-statistic map of the main effect for Composition as contrast of type `'1 1 1 1 1 1'`, reported at an uncorrected threshold of $P < 0.001$. Since activation during listening was not contrasted against some high-level baseline, but only against rest, this analysis is the most permissive, including all regions potentially (but not necessarily) involved in building up a semantic expectation of the upcoming visual array.

Estimation. Estimation refers to the visual phase in a trial when the display containing yellow and blue circles was presented after the subject listened to the sentence in the auditory Composition phase. In particular, the Estimation effect is operationalized as linear increase in BOLD signal with linear increase of the number of circles of the *TarCol.*

1. At the first level, the parametric increase of the BOLD signal with increasing number of circles of the respective color-of-mention was assessed separately for each quantifier on a trial-by-trial basis. The event-related GLM analysis for individual data sets involved 12 (2 × 6) orthogonal conditions, one for each color-of-mention (2) and quantifier (6). For each condition, a stick function (i.e., duration = 0, onset time = trial onset) was convolved with a canonical HRF and its first temporal derivative. Stick functions with duration = 0 were chosen in order to analyze the initial matching of the visual scenario with the mental representation generated in the auditory Composition phase before, and to address this process in the GLM independently of the actual duration of this matching process (this latter aspect relates to processing difficulty and is addressed with the analysis described next).
2. In order to model the parametric BOLD increase as a function of the number of circles (i.e., the data relevant for subsequent analysis), the percentage of circles in the *TarCol* was entered as monotonic, continuous regressor of interest into the model on a trial-by-trial basis separately for each condition. For subsequent ANOVA at the second level, we were interested in the parametric BOLD increase for each quantifier independent of the actual color (blue, yellow). Therefore, the beta weights for the two regressors (blue, yellow) for the same quantifier were collapsed pair-wise over colors, resulting in one average beta estimate per quantifier and subject. This was achieved using contrast vectors of the type `'1 1 0 0 0 0 0 0 0 0 0 0'` at the first level.
3. At the second level, these individual contrast images were submitted to a random-effects analysis, realized as a repeated-measures 1 × 6 ANOVA. In order to assess which brain regions uniformly responded with increasing activation to increasing numerosity, the T-statistic map for main contrast for Estimation was calculated as contrast of type `'1 1 1 1 1 1'`, reported at a threshold of $P < 0.001$ (uncorrected).

Comparison. Comparison is a process that calculates the numerical distance between the number of *TarCol* circles obtained during Estimation and its complement (the non-*TarCol* color). The matching between the outcome of the Comparison process and the meaning of the sentence will dictate the decision (e.g., *1/2* in the case of "*more-than-half of the circles are blue*") created during Composition on the other hand. As outlined in the Results section above, the processing of this numerical distance becomes the more difficult, the smaller this distance is. The difficulty, in turn, is represented in the RTs, with high RTs indicating high difficulty of Comparison[1].

Thus, the Comparison effects mainly reflect regions parametrically responding to increasing RT. Technically, identification of the Comparison was achieved as follows.

1. At the first level, a stick function (duration = 0) for each of the six conditions (one for each quantifier), was convolved with a canonical HRF and its first temporal derivative.
2. Additionally, for each trial, the RT for the *STVJ* task was entered as regressor of interest for the corresponding condition on a trial-by-trial basis. These regressors represent the parametric, non-monotonic increase of the BOLD response reflecting increase of processing difficulty. For subsequent ANOVA at the second level, the beta weights for each of the six parametric regressors of processing difficulty were contrasted against the implicit resting baseline in order to obtain T-statistic maps by calculating contrasts of the type `'1 0 0 0 0 0'`.
3. At the second level, these individual contrast images were submitted to a random-effects analysis, realized again as a repeated-measures 1 × 6 ANOVA. The T-statistic map for the main effect of the parametric RT regressor was computed as a contrast of type `'1 1 1 1 1 1'`, reported at an uncorrected threshold of $P < 0.001$.
4. In order to identify those voxels that were commonly involved in semantic Composition and responded to semantic processing difficulty (RT regressor), the T-statistic maps (at $P < 0.001$ uncorrected) obtained for the main effects for (auditory) Composition and (visual) Comparison part were submitted to a conjunction analysis in order to identify brain regions involved in a-modal semantic processing of quantities. We will henceforth refer to this conjunction analysis as the effect of Comparison.

Semantic structure. We used a set of quantifiers for the present study that we classified along the semantic dimension of Polarity, the absence or presence of linguistic negation. Negation is absent in positive quantifiers (*many, most, more-than-half*), but present in negative ones (*few, fewest, less-than-half*) quantifiers.

Effects of Semantic Structure – Polarity – can be examined in all three sets of parameters, i.e., Composition (the auditory sentence conditions), Estimation (the monotonic parameter of numerosity), and Comparison (the non-monotonic parameter of RTs). Reliability and robustness of effects of Semantic Polarity can thus be assumed if they are present in a given voxel not only for one but for all three parameter sets. Accordingly, the same contrast for Polarity was computed in each of the three parameter sets. Subsequently, the three T-statistic maps for Polarity effects (each at $P < 0.05$, $k = 100$ voxels) were submitted to a conjunction analysis revealing those voxels surviving this statistical threshold in all three contrasts. This conjunction analysis thus effects a conservative significance threshold of $P < 0.000125$.

Localization of effects with cytoarchitectonic probability maps

For the anatomical localization of the activations we used cytoarchitectonic probability maps, which are based on an observer-independent analysis of the cytoarchitecture in a sample of 10 post-mortem brains (Zilles et al., 2002; Schleicher et al., 2005). They provide information about the location and variability of cortical regions in standard MNI reference space. For the assignment of MNI coordinates to the cytoarchitectonically defined regions we used the SPM Anatomy Toolbox (Eickhoff et al., 2005) available with all published cytoarchitectonic probability maps and references from http://www.fz-juelich.de/inm/inm-1/spm_anatomy_toolbox

RESULTS

BEHAVIORAL DATA

The following behavioral results were found: First, participants' responses (truth-value judgments) presented a step-function along the "yes–no" axis (where "yes" indicates a sentence-scenario match, **Figure 2**), indicating high overall performance. For all six quantifier conditions, the Estimation parameter c (the number of circles in the Target Color) was highly correlated with mean response type (yes/no; $|r| > 0.8$, $P < 0.001$ in every instance, with quantifiers' polarity determining sign; **Table 4**). Second, mean RTs, when plotted against c in the images, formed non-monotonic functions, with maxima around the middle ($n = 25$, **Figure 2**). The correlation between the c and mean RT was low (not significant for four quantifier conditions, and significant but low (i.e., $r < 0.30$) for *more-than-half* and *many;* **Table 4**). Third, mean response type (yes/no) and RT were uncorrelated for three quantifier conditions; correlation for the other three conditions was significant, but low in every instance (**Table 4**).

BRAIN REGIONS ACTIVATED IN ESTIMATION AS COMPARED TO COMPARISON

Brain regions responsive to Estimation were identified via the monotonic parameter c, which varied incrementally ($c = 5$, 10,. . .40, 45) and was correlated with the BOLD fMRI signal. Effects of this parametric analysis (at $P_{\text{uncorr}} < 0.001$) were observed in both hemispheres, mainly in parietal and frontal regions including bilateral inferior parietal lobule (IPL), bilateral IPS, left SPL, left inferior frontal and right middle frontal gyrus (IFG; MFG), and the supplementary motor area (SMA; **Figure 3**, **Table 2** for a full list). The cytoarchitectonic probabilistic brain atlas (Amunts et al., 2004) implemented in the SPM Anatomy toolbox (Eickhoff et al., 2005) revealed that parietal effects overlapped with areas hIP1-3 of the IPS (Choi et al., 2006; Scheperjans et al., 2008), and areas PGa, PFm, PF of the IPL (Caspers et al., 2006). Frontal effects extended to Broca's region (areas 44 and 45; Amunts et al., 1999) and area 47 (Talairach daemon: http://www.nitrc.org/projects/tal-daemon; see also Amunts et al., 2010) in the IFG as well as premotor area 6 in the precentral gyrus (Geyer, 2003).

Regions responsive to Comparison were identified via the RT parameter. It varied non-monotonically and was made a regressor for the BOLD fMRI response Responses to Comparison

---

[1] In order to fully appreciate the effects of the monotonic regressor for Estimation and the non-monotonic RT regressor for Comparison, these two effects were assessed in two separate GLM analyses for each subject at the first level. Effects due to shared variance of the two regressors, if present, are identified as overlaps of the resulting T-statistic maps at the second level analyses (i.e., technically, as logical conjunctions).

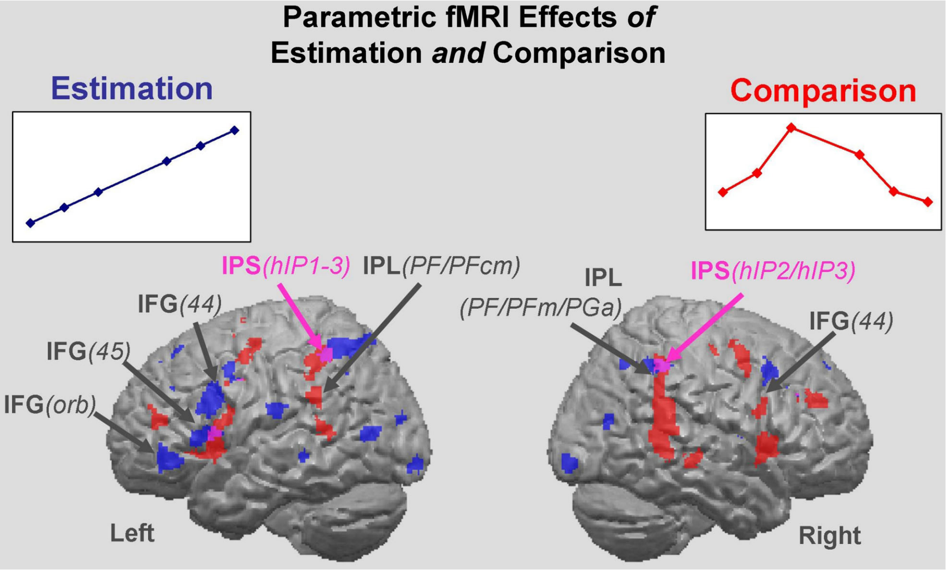


FIGURE 3 | Estimation and Comparison regions. Parametric effects for the Estimation of the numerosity of circles in target color (blue areas), and for the Comparison as manifest through RT (red areas). All effects are main effects over the six quantifiers. Estimation clusters: brain regions in which the BOLD signal intensity co-varies with the linearly increasing number of circles in target color (*F*-test). Comparison clusters: brain regions in which the BOLD signal intensity co-varies with RT – a reflection of processing difficulty. Regions commonly tapped by the Estimation and Comparison parameters are shown in purple (red plus blue). IPS, intraparietal sulcus; IFG, inferior frontal gyrus; IFG/orb, pars orbitalis of the IFG.

also involved a bilateral fronto-parietal set of regions. Parietal effects included bilateral IPL, bilateral IPS, and left SPL. Frontal regions comprised bilateral IFG and SMA. In addition, the putamen was activated (**Figure 3**; **Table 3** for a full list). Cytoarchitectonic probability maps revealed the involvement of frontal areas 44, 6, and 4a, and parietal areas hIP1-3 and PF, PFm, and PFcm.

Although both the Estimation effect and the Comparison effect recruited fronto-parietal regions, they overlapped only in part having centroids that were separate in each region (**Figure 3**). In the inferior parietal lobule, the Comparison effect was located anterior–inferior to the Estimation effect. Both effects involved areas in the IPS (hIP1-3) and inferior parietal lobule (PFm, PF). However, whereas Comparison uniquely involved area PFcm, Estimation recruited inferior parietal area PGa as well as superior parietal areas 7A and 7PC, allowing the cytoarchitectonic distinction between Estimation and Comparison regions within the parietal cortex. In the frontal cortex, a similar pattern of effects was observed with Estimation recruiting more anterior (and slightly inferior) aspects of the IFG, precentral gyrus, and SMA than Comparison. These were mostly distinct from the more posterior Comparison effects at the pre-defined significance threshold. Again, Estimation and Comparison conjointly recruited only some areas (i.e., 44 and 6). Interestingly, however, the Comparison effect also extended more posteriorly, covering parts of motor area 4a, whereas the Estimation effect reached more anteriorly into area 45.

SEMANTIC POLARITY CONTRAST

We compared the fMRI effect of Polarity (negative > positive quantifiers) in all three sets of parameters, i.e., for the non-monotonic Comparison parameter, for the monotonic Estimation parameter, and also for the BOLD signal in the auditory phase of the trial when the quantifier sentence was presented. Consistency of a Polarity effect in all three parameter sets was tested with a conjunction analysis revealing only voxels responding (at $P < 0.05$; extent $k = 100$ voxels) likewise in all three parameter sets (i.e., an overall significance level of $P < 0.000125$; see Supplementary Material for details). The most prominent effect was in the left inferior frontal gyrus (cytoarchitectonic area 45; **Figure 4**).

DISCUSSION

The separation between semantic Polarity, numerical Estimation, and numerical Comparison was achieved via decisions that participants made on visual scenarios in which numerosity (and subsequently proportion) was parameterized. These scenarios were presented in contrasting linguistic contexts. Though embedded in a tightly controlled design, the PPP was implemented in a rather naturalistic verification task: we verify sentences daily, in communicative acts that require us to answer a yes/no question. These can range from the most mundane topics (*Are you wearing a clean shirt?*) to complex ones that contain quantifiers (*Is every man in the room wearing a tie?*). We also engage in verification when we evaluate statements, just like in the PPP task (*Most of the paintings in this gallery are by Lucian Freud*). The choice of proportional quantifiers, whose calculation requires both Estimation and Comparison, helped in keeping the probing method constant while systematically varying both the numerosities and task difficulty so as to allow for the bi-parametric analysis through the PPP.

There are two novel aspects here: First, these data were obtained for the first time from the same set of subjects in a single study,

Table 2 | Main clusters in which signal increased as a function of parametric increase of stimation (number of circles of the mentioned color), with their cytoarchitectonic reference.

Cluster size (voxels)	Local maximum in macroanatomical structure	*x*	*y*	*z*	T_{max}	Percent of cluster volume in cytoarchitectonic area	
Cluster 1 (1930)	Left inferior parietal lobule (cluster extends into IPS)	−45	−53	57	5.30	32.2	hIP1
Cluster 2 (1685)	Left medial frontal gyrus	0	24	44	4.99	16.6	hIP3
Cluster 3 (1404)	Left precentral gyrus (cluster extends into IFG)	−39	5	33	6.77	43.2	Area 44
						16.8	Area 45
Cluster 4 (373)	Right middle frontal gyrus	39	14	38	4.46		
Cluster 5 (364)	Left inferior frontal gyrus	−48	20	8	5.18	56.9	Area 44
						15.1	Area 45
Cluster 6 (304)	Right supramarginal gyrus (cluster extends into right angular gyrus and IPS)	48	−42	45	4.76	48.4	PFm
						19.1	PF
						13.8	hIP2
Cluster 7 (291)	Left inferior frontal gyrus	−53	41	−8	6.02		
Cluster 8 (232)	Right angular gyrus	36	−60	45	3.94	27.6	PGa
						13.8	hIP3
Cluster 9 (223)	Left middle temporal gyrus	−51	−38	−8	4.21		
Cluster 10 (159)	Right inferior occipital gyrus	32	−95	−11	4.52	69.2	hOC3v
						26.4	Area 18
Cluster 11 (139)	Left inferior occipital gyrus	−26	−98	−11	4.18	48.9	hOC3v
						41.7	Area 18
Cluster 12 (82)	Right medial frontal gyrus	8	48	44	4.15		
Cluster 13 (74)	Left middle occipital gyrus	−32	−69	32	4.17		
Cluster 14 (58)		44	17	−15	3.70		
Cluster 15 (36)	Right middle orbital gyrus	27	41	−14	3.94		
Cluster 16 (36)	Right cerebellum	21	−86	−35	3.65		
Cluster 17 (26)	Left SMA	−15	3	65	3.51	96.2	Area 6
Cluster 18 (21)	Left superior frontal sulcus	−21	42	18			

References to cytoarchitectonic maps: area 2: Grefkes et al. (2001)*; areas 4a/4p:* Geyer et al. (1996)*; areas hIP3/7A/7M/7P/7PC:* Scheperjans et al. (2008)*; area 6:* Geyer (2003)*; areas 17/18:* Amunts et al. (2000)*; areas 44/45:* Amunts et al. (1999)*; areas PGa/PGp/PFm/PF/PFcm:* Caspers et al. (2006)*; areas hIp1/hIp2:* Choi et al. (2006)*; areas hOC3v/hOC4v:* Rottschy et al. (2007)*; areas Hipp(SUB)/Hipp(CA):* Amunts et al. (2005)*; area OP2:* Eickhoff et al. (2006)*. Further abbreviations:* T_{max}*, T value at local maximum. Cluster overlap with cytoarchitectonic areas is listed if it exceeds 10%.*

instead of merging together disparate studies of different scopes and quality. Second, and most importantly, it is the spatial relation of the semantic Polarity regions to the arithmetical parts, exposed through the parametric analyses of the *PPP*'s Proportion and RT parameters that served as proxies for Estimation and Comparison, respectively. Not only were these two neurocognitive components successfully separated in humans for the first time on the same data set (and in keeping with earlier findings – Piazza et al., 2004; Cohen Kadosh et al., 2005 – and sketches – Dehaene et al., 2003), but also, their spatial relation to activations during semantic analysis was demonstrated: Bilateral parietal regions were shown to be entrusted with numerical Estimation (via the *c* parameter), and adjacent regions on the opposite side of the IPS that support numerical Comparison (via the RT parameter, uncorrelated with the previous one). A new picture of a processing sequence along the "dorsal stream" seems to be emerging. The original notion of the dorsal stream was that of "where" pathway in visual perception, relevant for localizing objects in space. The concept of a functional processing pathway connecting occipito-temporal brain regions to the frontal cortex via the parietal lobe was supported by an anatomical connectivity study (Caspers et al., 2011). In this study, analogies between white matter tracts in humans and different kinds of monkeys were established. Interestingly, though, the existence of such anatomical connectivity does not imply information about the role of the dorsal stream in cognitive processing. A recent hypothesis (Friederici et al., 2012) relates the maturation of the dorsal stream to linguistic, and in particular syntactic, development from childhood to adulthood. Parallel to that notion, research in the field of numerosity processing (Dehaene et al., 2003) suggested a functional parcelation within the parietal aspect of the dorsal stream: (1) The angular gyrus is regarded as a region for manipulation of verbal number representations. (2) The horizontal aspect of the IPS is thought to house the mental number line. (3) The superior parietal cortex, finally, might relate to attentional processing when orientation along the mental number line is concerned.

Our study relates to the hypothesis by Dehaene et al. (2003) in that it provides an empirical basis for a bi- (or even tri-)partition of the parietal aspect of the dorsal stream, which was obtained

Table 3 | Conjunction analysis representing brain areas involved in comparison (abstract representation of numerosity).

Cluster size (voxels)	Local maximum in macroanatomical structure	*x*	*y*	*z*	$T_{\max}$	Percent of cluster volume in cytoarchitectonic area	
Cluster 1 (6059)	Left SMA	−5	6	53	7.72	28.0	Area 6
						18.3	Area 6
Cluster 2 (4124)	Right putamen	20	14	−5	7.36		
Cluster 3 (3964)	Left putamen (cluster extends into IFG)	−18	17	−8	6.44	10.4	Area 44
Cluster 4 (2163)	Right inferior parietal lobule	63	−42	17	7.16	24.6	PFm
						13.2	PF
Cluster 5 (831)	Left inferior parietal lobule	−48	−47	50	4.29	23.0	hIP2
	(cluster extends into IPS)					20.6	hIP1
						19.6	hIP3
						18.9	PF
Cluster 6 (571)	Right precentral gyrus	41	0	50	4.53	29.1	Area 6
Cluster 7 (317)	Right middle frontal gyrus	38	42	27	4.14		
Cluster 8 (184)	Left supramarginal gyrus	−57	−41	27	4.38	77.2	PF
	(cluster extends into						
	left angular gyrus)					22.8	PFcm
Cluster 9 (177)	Left middle temporal gyrus	−56	−50	9	4.27		
Cluster 10 (163)	Left inferior frontal gyrus	−41	42	15	4.27		
Cluster 11 (119)	Left calcarine sulcus	−23	−71	8	4.21	41.2	Area 17
Cluster 12 (110)	Right inferior frontal gyrus	44	8	24	3.88	54.5	Area 44
Cluster 13 (92)	Right inferior frontal gyrus	56	11	20	4.01	98.9	Area 44
Cluster 14 (90)	Left thalamus	−8	−8	3	3.68		

Parametric increase in brain activation as a function of processing difficulty (reaction time) is intersected with brain activation while listening to sentences containing a quantifier word and building up a semantic representation of the upcoming visual display. For further details see legend of ***Table 2****.*

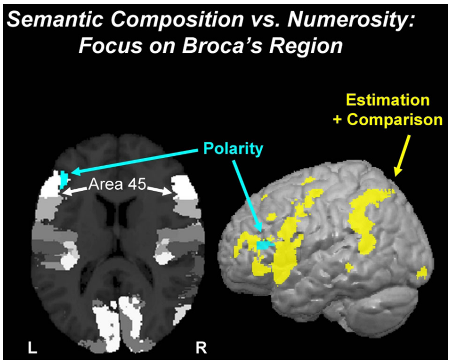


FIGURE 4 | Semantics of quantification: polarity. The effect of Polarity (negative vs. positive quantifiers; turquoise blob) was observed in Broca's region in the left inferior frontal gyrus. The effect, which is clearly non-numeric but semantic in nature, overlaps with cytoarchitectonic area 45 (white). The surface rendering shows that this Polarity effect is predominant in the left inferior frontal cortex, whereas the numerosity network (yellow) comprising both areas for Estimation and Comparison (from **Figure 3**) taps widely into a fronto-parietal network.

with the same set of participants in the same experimental session. Regions for numerical Estimation and Comparison were found, which anatomically correspond to the angular and SPL regions referred to by Dehaene et al. (2003) and which are indeed separated by the IPS. Moreover, while the present numerosity results are consistent with similar ones from the macaque (Nieder and Miller, 2003; Roitman et al., 2007; Piazza and Izard, 2009; Caspers et al., 2011), our data, obtained from humans instead of animals, for the first time establish a clear connection to linguistic analysis of quantity, which builds upon and extends previous findings: Whereas a fronto-parietal network emerges for Estimation and Comparison, linguistic–semantic analysis is most prominent in the left inferior frontal cortex. Taken together, our data characterize processing along the dorsal stream from inferior parietal to inferior frontal regions when numerical processing is linked to linguistic–semantic processing. With respect to the hypothesis by Friederici et al. (2012), the present study may suggest that the dorsal stream is indeed involved in linguistic processing. However, over and above its supposed role for the emergence of syntactic abilities, our data suggest its additional importance for the transformation from visual percepts of numerosities into cognitive and linguistic formats. Interestingly, this account parallels the syntax hypothesis in so far as it stresses the difference between monkey and man with respect to linguistic abilities.

We thus found a potential human/monkey homology for successive steps of numerosity assessment, but moreover, established a direct connection to language. At the same time, we provided evidence for the neural modularity of language and arithmetic, revealed through a task that matches linguistic representations with numerical ones. Such matching has been used before to distinguish numbers from numerosities (Cohen and Dehaene,

Table 4 | Correlations between the Estimation parameter c, average response rate in the STVJ task, and corresponding RT for each quantifier.

Quantifier	*r*	*P*
Correlation of Estimation parameter *c* with average response		
Few	−0.825	<0.001
Fewest	−0.849	<0.001
Less-than-half	−0.822	<0.001
Many	0.869	<0.001
More-than-half	0.884	<0.001
Most	0.869	<0.001
Correlation of Estimation parameter *c* with RT		
Few	0.043	0.531
Fewest	0.108	0.114
Less-than-half	−0.096	0.158
Many	0.263	<0.001
More-than-half	0.184	0.007
Most	0.097	0.157
Correlation of Average response with RT		
Few	−0.002	0.972
Fewest	−0.043	0.531
Less-than-half	0.135	0.047
Many	0.178	0.009
More-than-half	0.198	0.003
Most	0.126	0.064

2001), and in the context of previous work on quantification (McMillan et al., 2005; Troiani et al., 2009). Our contribution, supporting the modularity of language and arithmetic, stems from the fact that the task we used recruits not only lexical or sentential, but also complex compositional, resources, thereby better approximating the much discussed Chomsky's "mental organ for language." Whereas parietal regions were involved only in arithmetical but not semantic analysis, frontal regions, by contrast, appear to be involved in both. This pattern suggests that the left inferior frontal cortex contributes to the evaluation of the numerical representations provided in the parietal regions. This is consistent with the fact that in the macaque brain, number-selective neurons (Estimation) in inferior frontal cortex respond later than number-sensitive (Comparison) parietal neurons (Piazza and Izard, 2009). This temporal delay suggests an order, whereby frontal modules evaluate the output from parietal modules. In humans, this frontal contribution relates to semantic structure and may thus additionally be linked to linguistic processing.

Interestingly, functional division between Estimation and Comparison similar to that in parietal cortex was also found in both frontal lobes. Likewise, the semantic Polarity analysis was prominent in the left frontal lobe in area 45. Evidence for frontal involvement in numerical cognition has been available (Piazza et al., 2007), but the present results are surprising, as they seem to suggest commonalities in the functional organization between the inferior frontal cortex and the peri-IPS regions. McMillan et al. (2005), who studied the relation between natural language quantifiers and numerosity, have argued for a functional distinction, such that frontal regions house working memory, and parietal ones are entrusted with knowledge of numbers. The present study demonstrated the involvement of both frontal and parietal cortex areas likewise in response to Estimation and Comparison, even thought these processes may require working memory resources to different extents. Thus, the present data provide no evidence in support of the hypothesis by McMillan et al. (2005). Moreover, their distinction that was made with respect to the underlying anatomy, however, is possibly not fine-grained enough. Given the significant architectonic difference between areas 44 and 45 (involved in the Comparison network) on the one hand, and Brodmann's area 47 (involved in the Estimation network) on the other hand (Amunts et al., 2010), it seems difficult to maintain that all frontal clusters uniformly support the same function, be it working memory or other, during numeric cognition. Certainly, there is a wealth of evidence for structural and functional connectivity between anterior IPS areas and ventral–posterior inferior frontal cortex (Uddin et al., 2010; Caspers et al., 2011; Mars et al., 2011). But whether these really involve working memory loops (Vigneau et al., 2011), or rather a house a sequential evaluation algorithm for the evaluation of previously estimated numerosities, as proposed here on the basis of a well-controlled experiment, will be the objectives of future research of the underlying structural and functional connectivity. The significance of this work, then, is in exposing a rich array of regions that provide a rare window into the intriguingly complex neural system that handles the flow of information between neural substrates for linguistic and numerical cognition.

ACKNOWLEDGMENTS

We are grateful to Danny Fox, Israel Nelken, Yonatan Loewenstein, Lew Shapiro, and Michael Wagner for their incisive and helpful comments. Furthermore, we wish to thank Peter Pieperhoff, Berhard Schwarz, and Andrea Santi for advice on methodology. Partial support for this project was provided by an Alexander von Humboldt Foundation Research Award (Yosef Grodzinsky) and by NIH (grant #00094), as well as SSHRC (standard grant #410-2009-0431) and Canada Research Chairs (Yosef Grodzinsky).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Evolutionary_Neuroscience/10.3389/fnevo.2012.00004/abstract

Supplement S1 | Properties of the images and how they were created.

Supplement S2 | The mapping between the current PPP and standard numerosity experiments.

Supplement S3 | The meaning of the proportional quantifiers used.

Supplement S4 | Correlations between the proportion parameter *c*, average response rate in the STVJ task, and corresponding RT for each quantifier.

REFERENCES

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N. J., Habel, U., Schneider, F., and Zilles, K. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat. Embryol.* 210, 343–352.

Amunts, K., Lenzen, M., Friederici, A. D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., and Zilles, K. (2010). Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol.* 8, e1000489. doi:10.1371/journal.pbio.1000489

Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., and Zilles, K. (2000). Brodmann's areas 17 and 18 brought into stereotaxic space-where and how variable? *Neuroimage* 11, 66–84.

Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H. B. M., and Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412, 319–341.

Amunts, K., Weiss, P. H., Mohlberg, H., Pieperhoff, P., Eickhoff, S., Gurd, J. M., Marshall, J. C., Shah, N. J., Fink, G. R., and Zilles, K. (2004). Analysis of neural mechanisms underlying verbal fluency in cytoarchitectonically defined stereotaxic space – The roles of Brodmann areas 44 and 45. *Neuroimage* 22, 42–56.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16, 4207–4221.

Brannon, E. M. (2005). The independence of language and mathematical reasoning. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3177–3178.

Butterworth, B. (1999). A head for figures. *Science* 284, 928–929.

Cantlon, J. F., and Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biol.* 5, e328. doi:10.1371/journal.pbio.0050328

Caspers, S., Eickhoff, S. B., Rick, T., von Kapri, A., Kuhlen, T., Huang, R., Shah, N. J., and Zilles, K. (2011). Probabilistic fibre tract analysis of cytoarchitectonically defined human inferior parietal lobule areas reveals similarities to macaques. *Neuroimage* 58, 362–380.

Caspers, S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., and Zilles, K. (2006). The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual. *Neuroimage* 33, 430–448.

Changeux, J. P., and Connes, A. (1995). *Conversations on Mind, Matter, and Mathematics*. Princeton, NJ: Princeton University Press.

Choi, H. J., Zilles, K., Mohlberg, H., Schleicher, A., Fink, G. R., Armstrong, E., and Amunts, K. (2006). Cytoarchitectonic identification and probabilistic mapping of two distinct areas within the anterior ventral bank of the human intraparietal sulcus. *J. Comp. Neurol.* 495, 53–69.

Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.

Cohen, L., and Dehaene, S. (2000). Calculating without reading: unsuspected residual abilities in pure alexia. *Cogn. Neuropsychol.* 17, 563–583.

Cohen, L., and Dehaene, S. (2001). Occam's razor is not a swiss-army knife: a reply to pillon and pesenti. *Cogn. Neuropsychol.* 18, 285–288.

Cohen Kadosh, R., Henik, A., Rubinsten, O., Mohr, H., Dori, H., van de Ven, V., Zorzi, M., Hendler, T., Goebel, R., and Linden, D. E. (2005). Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia* 43, 1238–1248.

Dale, A. M., and Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Hum. Brain Mapp.* 5, 329–340.

Dehaene, S., and Changeux, J. P. (1993). Development of elementary numerical abilities: a neuronal model. *J. Cogn. Neurosci.* 5, 390–407.

Dehaene, S., and Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33, 219–250.

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506.

Diesing, M. (1992). *Indefinites*. Cambridge, MA: MIT Press.

Eickhoff, S. B., Amunts, K., Mohlberg, H., and Zilles, K. (2006). The human parietal operculum. II. stereotaxic maps and correlation with functional imaging results. *Cereb. Cortex* 16, 268–279.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., and Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335.

Friederici, A. D., Oberecker, R., and Brauer, J. (2012). Neurophysiological preconditions of syntax acquisition. *Psychol. Res.* 76, 204–211.

Gelman, R., and Butterworth, B. (2005). Number and language: how are they related? *Trends Cogn. Sci.* 9, 6–10.

Geyer, S. (2003). *The Microstructural Border Between the Motor and the Cognitive Domain in the Human Cerebral Cortex*. Wien: Springer.

Geyer, S., Ledberg, A., Schleicher, A., Kinomura, S., Schormann, T., Bürgel, U., Klingberg, T., Larsson, J., Zilles, K., and Roland, P. E. (1996). Two different areas within the primary motor cortex of man. *Nature* 382, 805–807.

Grefkes, C., Geyer, S., Schormann, T., Roland, P., and Zilles, K. (2001). Human somatosensory area 2: observer-independent cytoarchitectonic mapping, interindividual variability, and population map. *Neuroimage* 14, 617–631.

Mars, R. B., Jbabdi, S., Sallet, J., O'Reilly, J. X., Croxson, P. L., Olivier, E., Noonan, M. P., Bergmann, C., Mitchell, A. S., Baxter, M. G., Behrens, T. E., Johansen-Berg, H., Tomassini, V., Miller, K. L., and Rushworth, M. F. (2011). Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *J. Neurosci.* 31, 4087–4100.

McMillan, C. T., Clark, R., Moore, P., Devita, C., and Grossman, M. (2005). Neural basis for generalized quantifier comprehension. *Neuropsychologia* 43, 1729–1737.

Meck, W. H., and Church, R. M. (1983). A mode control model of counting and timing processes. *J. Exp. Psychol. Anim. Behav. Process* 9, 320–334.

Milsark, G. (1977). Toward an explanation of certain peculiarities of the existential construction in english. *Ling. Analysis* 3, 1–29.

Nieder, A., and Miller, E. K. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 37, 149–157.

Piazza, M., and Izard, V. (2009). How humans count: numerosity and the parietal cortex. *Neuroscientist* 15, 261–273.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* 44, 547–555.

Piazza, M., Pinel, P., Le Bihan, D., and Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron* 53, 293–305.

Roitman, J. D., Brannon, E. M., and Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS Biol.* 5, e208. doi:10.1371/journal.pbio.0050208

Rottschy, C., Eickhoff, S. B., Schleicher, A., Mohlberg, H., Kujovic, M., Zilles, K., and Amunts, K. (2007). Ventral visual cortex in humans: cytoarchitectonic mapping of two extrastriate areas. *Hum. Brain Mapp.* 28, 1045–1059.

Scheperjans, F., Eickhoff, S. B., Hömke, L., Mohlberg, H., Hermann, K., Amunts, K., and Zilles, K. (2008). Probabilistic maps, morphometry and variability of cytoarchitectonic areas in the human superior parietal cortex. *Cereb. Cortex* 18, 2141–2157.

Schleicher, A., Palomero-Gallagher, N., Morosan, P., Eickhoff, S. B., Kowalski, T., de Vos, K., Amunts, K., and Zilles, K. (2005). Quantitative architectural analysis: a new approach to cortical mapping. *Anat. Embryol.* 210, 373–386.

Troiani, V., Peelle, J. E., Clark, R., and Grossman, M. (2009). Is it logical to count on quantifiers? Dissociable neural networks underlying numerical and logical quantifiers. *Neuropsychologia* 47, 104–111.

Uddin, L. Q., Supekar, K., Amin, H., Rykhlevskaia, E., Nguyen, D. A., Greicius, M. D., and Menon, V. (2010). Dissociable connectivity within human angular gyrus and intraparietal sulcus: evidence from functional and structural connectivity. *Cereb. Cortex* 20, 2636–2646.

Varley, R. A., Klessinger, N. J., Romanowski, C. A., and Siegal, M. (2005). Agrammatic but numerate. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3519–3524.

Verguts, T., and Fias, W. (2004). Representation of number in animals and humans: a neural model. *J. Cogn. Neurosci.* 16, 1493–1504.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B., and Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? insights from a meta-analysis. *Neuroimage* 54, 577–593.

Zilles, K., Schleicher, A., Palomero-Gallagher, N., and Amunts, K. (2002). "Quantitative analysis of cyto- and receptor architecture of the human brain" in *Brain Mapping, The Methods*, eds J. Mazziotta and A. Toga (San Diego: Academic Press), 573–602.

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: 23 November 2011; paper pending published: 29 December 2011; accepted: 05 March 2012; published online: 21 March 2012.

Citation: Heim S, Amunts K, Drai D, Eickhoff SB, Hautvast S and Grodzinsky Y (2012) The language–number interface in the brain: a complex parametric study of quantifiers and quantities. Front. Evol. Neurosci. ***4****:4. doi: 10.3389/fnevo.2012.00004*

frontiers in
EVOLUTIONARY NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 12 June 2012
doi: 10.3389/fnevo.2012.00008

Learning to read aligns visual analytical skills with grapheme-phoneme mapping: evidence from illiterates

***Thomas Lachmann*[1]**, *Gunjan Khera*[1]*, *Narayanan Srinivasan*[2] *and Cees van Leeuwen*[3]

[1] *Center for Cognitive Science, Cognitive and Developmental Psychology Unit, University of Kaiserslautern, Kaiserslautern, Germany*
[2] *Centre of Behavioural and Cognitive Sciences, University of Allahabad, Allahabad, India*
[3] *KU Leuven, Experimental Psychology Research Unit, Laboratory for Perceptual Dynamics, Leuven, Belgium*

Edited by:
Michael Petrides, Montreal Neurological Institute and Hospital, Canada

Reviewed by:
Paul M. Nealen, Indiana University of PA, USA
Rosana M. Tristão, University of London, UK

****Correspondence:***
Thomas Lachmann, Center for Cognitive Science, Cognitive and Developmental Psychology Unit, University of Kaiserslautern, E.-Schroedinger-Str. 57, 67663 Kaiserslautern, Germany.
e-mail: lachmann@rhrk.uni-kl.de

Learning to read puts evolutionary established speech and visual object recognition functions to novel use. As we previously showed, this leads to particular rearrangements and differentiations in these functions, for instance the habitual preference for holistic perceptual organization in visual object recognition and its suppression in perceiving letters. We performed the experiment in which the differentiation between holistic non-letter processing and analytic letter processing in literates was originally shown (van Leeuwen and Lachmann, 2004) with illiterate adults. The original differentiation is absent in illiterates; they uniformly showed analytic perception for both letters and non-letters. The result implies that analytic visual perception is not a secondary development resulting from learning to read but, rather, a primary mode of perceptual organization on a par with holistic perception.

Keywords: letter recognition, object recognition, literacy, reading acquisition, dyslexia, flanker, illiterate participants, grapheme-phoneme conversion

Reading is a secondary process; learning to read depends on functional coordination of two established skills: visual object perception and the ability to use spoken language. Whereas, the former emerges at a much earlier stage than the second, both are well entrenched in human evolution. In learning to read, these skills are (1) recruited, (2) modified, (3) coordinated, and finally the coordinated process is (4) automated, in order to enable skillful, fluent reading (Lachmann, 2002). As a result, letters are detected and processed automatically in a cross-modal fashion (Blomert, 2011). To establish such a cross-modal representation requires long training, possibly 3–4 years of practice. Suboptimal automatization of this functional coordination may lead to reading disability (Lachmann et al., 2009; Blomert, 2011).

When successful, as a result of reading practice a differentiation in perceptual processing emerges (van Leeuwen and Lachmann, 2004; Burgund et al., 2006, 2009; Pegado et al., 2011). Whereas children whose reading skills are not yet fully automated process letters and similar geometric shapes in the same, holistic fashion: surrounding irrelevant visual information is uniformly assimilated with target letters and shapes in the early stages of visual feature binding, adult skilled readers, while still processing the shapes holistically, treat letters analytically: surrounding irrelevant visual context is ignored (Lachmann and van Leeuwen, 2004, 2008a) or even actively suppressed (van Leeuwen and Lachmann, 2004).

As readers acquire the differentiation in letter versus non-letter processing, the question could be asked: what perceptual skills do they lose (Serniclaes et al., 2005; Dehaene and Cohen, 2007; Lachmann and van Leeuwen, 2007; Blomert, 2011)? There is evidence that might be taken to suggest that normal reading involves losing the ability to process letters holistically. With letters and dot-patterns in a non-lexical *same–different* task, symmetry in dot patterns benefitted both normal reading children and their dyslexic peers, whereas symmetry in letters benefitted the dyslexics but not the normal readers. As a result, normally reading children were *slower* for letters in this task than dyslexics (Lachmann and van Leeuwen, 2007). On the other hand, skills seemingly lost while in transition to fluent reading might resurface later when reading is fully integrated into our system. The contrasting strategies of analytic processing for letters versus holistic processing for shapes in a classification task gave way to uniformly holistic processing, once the task required that (van Leeuwen and Lachmann, 2004). Thus, literate adults cannot be said to have lost the ability to process letters holistically (see also Schwarzer et al., 2010). Rather, they seem to refrain from it habitually.

Perhaps we are not losing perceptual skills while learning to read, but are acquiring habits that sometimes lead to suboptimal performance on specific tasks (as in Lachmann and van Leeuwen, 2007). We may consider whether the specific perceptual strategies we *gain* from learning to read are likewise habitual preferences rather than newly acquired perceptual skills. Our previous research has remained inconclusive about whether prior to achieving letter-specific analytic processing, children are incapable of analytic processing, and hence this skill emerges from learning to read or, alternatively, whether this skill is already present, and is recruited for processing letters. On the one hand, children early in this process show uniformly holistic processing for letters and non-letters (Lachmann and van Leeuwen, 2008a). But on the other hand, certain subgroups of dyslexics in this study showed evidence of processing letters analytically, albeit with extreme effort. This might well be a particular consequence of the training these children have been receiving in school. However, from these results it cannot be determined whether children have learned analytic visual processing from learning to read.

To properly investigate the question what is lost or gained from learning to read (Dehaene and Cohen, 2007), we studied a group of adults who never learned to read. We compared these to a group of adult skilled readers, using the classification task in which the differentiation in processing between letters and non-letters was originally found (van Leeuwen and Lachmann, 2004). In this task, target letters and non-letters were shown either isolated or surrounded by a task-irrelevant shape (see **Figure 1**, for an example). Non-letters were classified faster if the target and its surrounding were form-congruent as compared to when they differed in shape, i.e., when both were form-incongruent. This is an example of the well-known congruence effect (Pomerantz and Pristach, 1989; Bavelier et al., 2000); and can be related to early and mid-level visual perception (van Leeuwen and Bakker, 1995; Boenke et al., 2009). The congruence effect indicates, in an operationally specific sense, holistic perceptual grouping: the surrounding visual information is bound to the target, and is processed faster if both call for the same response. For letter targets, however, the opposite result was found: letters were categorized faster when surrounded by an incongruent non-target than when the non-target was congruent, i.e., a negative congruence effect.

The observed dissociation was considered a product of a specific analytic strategy optimized during learning to read in order to guarantee a rapid grapheme-phoneme mapping. Binding irrelevant visual information from the surrounding would disturb this fast mapping. The visual structure of the surrounding shape is therefore suppressed (Lachmann, 2002). Doing so is more difficult for congruent than for incongruent items, resulting in negative congruence effects (Briand, 1994; van Leeuwen and Bakker, 1995; Bavelier et al., 2000). Thus, we may conclude that in the early or intermediate stages of visual perception, skilled readers process letters using a unique analytic encoding strategy whereas for processing shapes a holistic processing strategy is applied.

For adults who have never learned to read and are unfamiliar with the alphabet, we should not expect such dissociation. Since they are not able to differentiate letters from non-letters, they will process both letters and non-letter shapes with one and the same strategy (e.g., Dehaene et al., 2010b). The question is: will that be a holistic or an analytic strategy? If the former, we may conclude that the analytic strategy is a secondary processing strategy; if it is an analytic strategy we should conclude that both analytic and holistic processing are intrinsic, primary strategies of the visual system.

METHOD

PARTICIPANTS

Fifty-eight adults participated in this experiment. All of them reported normal vision and hearing. The participants were informed about the reason for this investigation, that participation is voluntary and that they are free to withdraw their participation at any time while keeping their reward. The ethical committees of both the University of Allahabad and the University of Kaiserslautern gave their approval to the experiment.

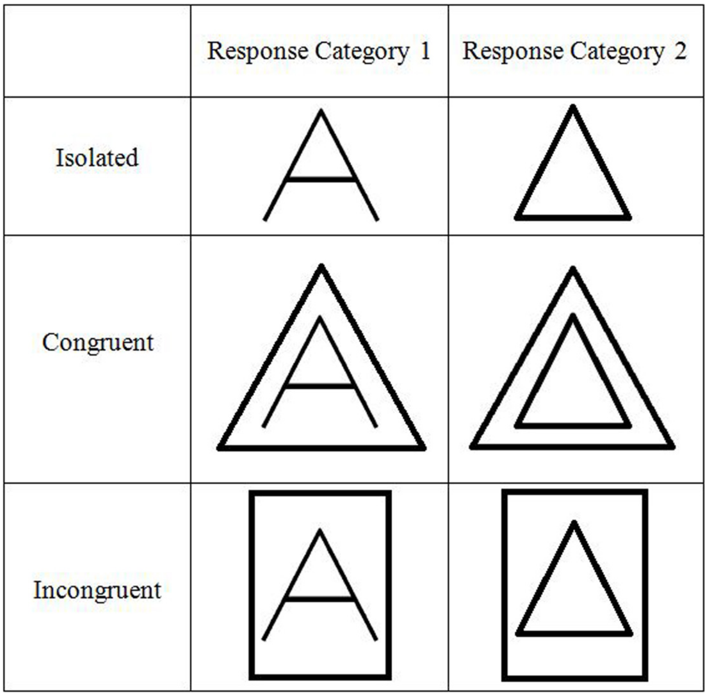


FIGURE 1 | Example for a letter target (left column) and a similar shape target (right column) as presented isolated (first line), surrounded by a form-congruent (second line) or a form-incongruent non-target in van Leeuwen and Lachmann (2004).

The control group consisted of 26 Indian students (seven females), aged between 22 and 29 years ($M = 26$ years, $\text{SD} = 1.97$) from the University of Kaiserslautern, Germany. All of them were able to read fluently and to write in English. They gave their consent to the participation in written form and received 5 Euros (equivalent of 6.5 USD) for performing in this experiment.

The experimental group consisted of 32 Indian illiterates (18 females), aged between 20 and 31 years ($M = 28$ years, $\text{SD} = 3.71$), from a suburb of Jasra village (Gadaiya Kalan, approximately 450 inhabitants, agriculture area), located about 35 km south of the city of Allahabad (Uttar Pradesh). These participants have been living in the village since they were born. Because there was no public school close by when they were children, none of them attended schooling. All participants are involved in farming and sell their products to external dealers coming to the village. There is no library in the village, nor is there any access to newspapers. Participants of this group received 150 Rs (equivalent to about 3 USD) for completing the experiment. This amount is about equivalent to an average daily wage and covers their loss in income for the day, when participating in the experiment. All these participants reported that they are not able to read neither English nor Hindi, do not speak English, and are not familiar with Latin alphabets. These participants gave their consent to the participation in spoken form.

Prior to the experiment illiterate participants' familiarity with the alphabet was determined by using a simple paper and pencil letter identification task, which included 72 letters and non-letter items, aligned in six rows. Each row contained four letters and eight non-letters, so in total 24 different letters and 48 different non-letters were displayed. Non-letters were constructed in a way that they consisted of the same segments as, but do not look too similar to Latin nor to Hindi letters (see **Figure 2**). Note, that non-letters used in this pre-task were not the geometrical shapes used in the experiment. Participants were instructed to mark those four items in each row of which they think they were letters. They were informed that payment does not depend on this performance. In case they were unable to decide (typical statement: "this all looks English to me"), they were instructed to guess.

The mean number of correct responses in the questionnaire was 10.35 ($\text{SD} = 4.4$). Guessing rate was calculated on the basis of the hyper-geometric distribution [$X \sim Hyp(n, H, F)$]; with n is the number of guesses, H is the number of possible hits, and F the number of possible false alarms. Accordingly, the expected value of the guess rate is $E(X) = n \times [H/(H + F)] = 1.33$ correct responses per row. For six rows the total $E(X)$ is $6 \times E(X) = 8$. Taking $E(X) \pm 1\ \text{SD}(X)$ as criterion, with $\text{SD}(X) = 4.82$, we accepted all participants of this group who identified between 0.77 and 15.23 letters correctly. Alternatively, guessing rate may be calculated on the basis of Tschebyscheff-inequality and the Gaussian, with the criterion $p < 0.5\%$, which would have resulted in a range of 0–17.44 hits. However, we adopted the first criterion as being more conservative. On this basis, one participant, who had identified zero letters correctly, was excluded from further analyses. Another participant did not finish the test and left the setting. The rest of the participants constitute the sample described above.

MATERIAL

As in van Leeuwen and Lachmann (2004, Experiment 4), there were 24 unique stimuli. They consisted of targets: the four capital letters A, H, L, C; and the four geometrical shapes square, triangle, rectangle, circle, each of which was shown either in isolation or surrounded by a congruent or incongruent non-target geometrical shape which were slightly enlarged versions of the triangle, square, rectangle, and circle targets. For instance, A in isolation, A surrounded by a triangle (congruent condition), A surrounded by a rectangle (incongruent condition).

The stimuli were scaled to an imaginary 50 × 50 mm matrix, and the surrounding shapes to an imaginary 80 × 80 mm matrix. Stimuli were presented in black ($0.29\ \text{cd/m}^2$) on a Laptop screen set to white ($27.3\ \text{cd/m}^2$) at about 50 cm distance, resulting in a visual angle of 3° without, and 4.5° with surrounding.

PROCEDURE

Literate participants performed in a laboratory room at the university. For the illiterate participants we rented a room in a private house in the village. The room was prepared in such a way that conditions were as similar to the laboratory as circumstances allowed. Electricity was guaranteed by using a mobile generator. The same portable computer was used for both groups. Participants of both groups were seated comfortably in the dimly lit environment. There was no head fixation.

The illiterate participants reported to have never used a computer before. Therefore, before instruction, they were familiarized with the computer and with the choice reaction task. They were presented with a set of short practice sessions, in which green and red dots were presented at the screen. In the first session participants were instructed to press a response key whenever a dot appears at the screen (simple reaction). In the second session a key press was required only if a red dot appeared (go/no-go-task) and, finally, in the third session the choice reaction task was trained: for red dots one key was to be pressed and one for green dots another. Prior to the experiment, literate and illiterate participants performed 12 warm up trials with the same task as used in the experiment. During practice and warming up sessions, feedback was given by presenting a yellow smiley after correct responses and a black cross after wrong responses.

Six different subsets of stimuli were presented to participants in a counterbalanced fashion (as in the original experiment,

FIGURE 2 | The fourth out of six rows used in the paper and pencil letter identification test.

see **Figure 2**). For each individual, the stimuli were restricted to two letters and two shapes. Letters and shapes were pairwise similar, e.g., an *A* and a triangle. They were assigned in a counterbalanced manner to two different response categories: for instance, Category 1 was an "*A* or a Rectangle" versus Category 2 was a "*L* or a Triangle" (Selection 1 in **Figure 2**). Only one stimulus was shown at a time, either with or without surrounding. The task was to decide, as fast as and accurate as possible, whether the shown target belongs to Response category 1 or 2, ignoring the surrounding if it appears. No feedback was given. Note that letters and shapes that are similar to each other in shape, such as the *A* and the triangle, were always assigned to different response categories (see example in **Figure 3**, Selection 1–3). Thus, in order to solve the task, phonological coding of the letters would be useful to distinguish between response categories, but is not necessary. In skilled readers this design was found to implicitly trigger a distinction between letters and shapes (van Leeuwen and Lachmann, 2004).

A speeded choice reaction task was required according to the response categories displayed in **Figure 3**, by pressing either the left or the right button marked on the keyboard of a portable computer. The four stimuli were shown to the participant prior to the experiment along with the assignment of stimuli to response buttons, e.g., if A or square press left button and if L or triangle, push right button. It was emphasized that the surrounding, if it occurs, is not relevant for the task and should be ignored.

Altogether each individual performed on 720 trials: 4 targets × 3 conditions (isolated, congruent surrounding, incongruent surrounding), with 60 repeated measures. Trials were randomized, having 12 breaks in between. For illiterates the experiment took about one hour including the letter test, instruction and practice; for controls it took about half an hour.

RESULTS

Reaction times (RT) for correct responses and error rates were analyzed after rejecting outliers, which were 0.5% for the criterion $RT < 145$ ms, and 0.3% for the individual criterion $RT > \mu_n + 6\sigma_n$. (μ_n = individual mean RT; σ_n = individual mean SD).

For illiterates mean error rate was 3.3% and ranged from 0.4 to 15%, three participants showed an error rate higher than 10%. Mean RT was 980 ms (SD = 455 ms).

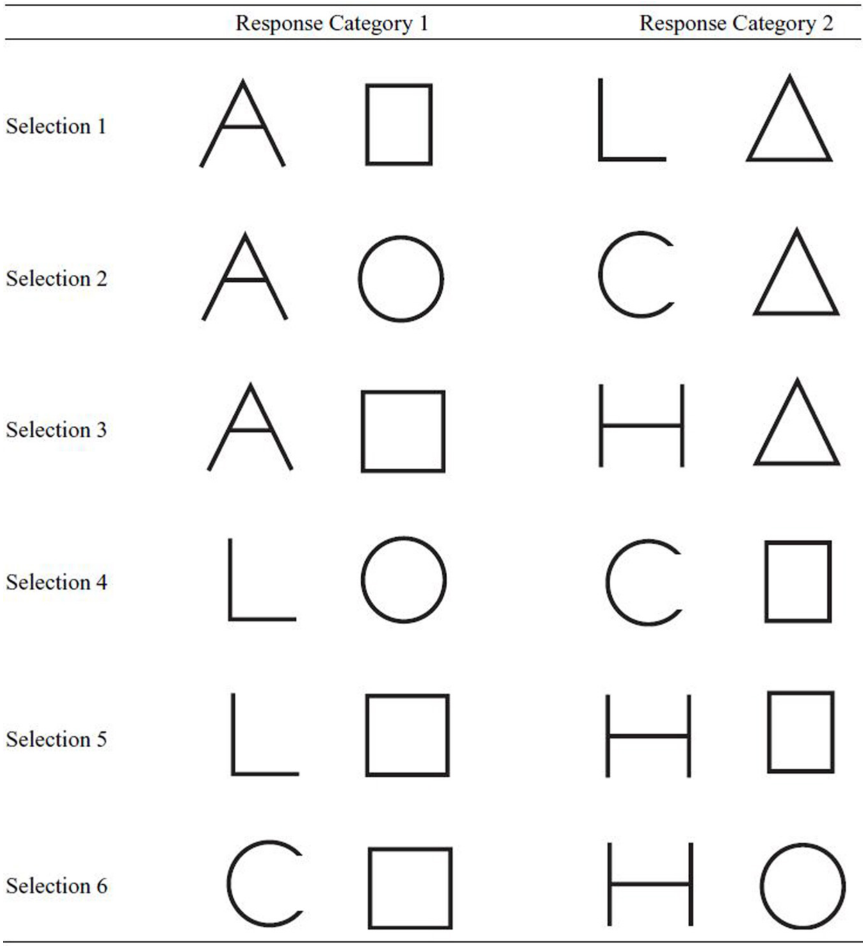


FIGURE 3 | Six selections of stimuli used in the experiment for individual participants. Further explanations in the text.

For literates, mean error rate was 3.1%, significantly below the 7.3% in the original study with German adults, $F_{(1,\ 49)} = 10.1$, $p < 0.01$, and ranged from 0.1 to 10%. The error rate did not differ from that of illiterate participants of the present study. Mean RT was 454 ms (SD = 186), which did not differ from German adults in the original study (477 ms, SD = 126) but is significantly lower, $F_{(1,\ 56)} = 111.2$, $p < 0.001$, than that in illiterates, whose RTs are more than double. Only two literate participants had average RTs over 600 ms and only one had a higher mean RT than the fastest participant of the illiterate group.

Because there was no speed-accuracy trade-off in evidence, i.e., individual mean error rates and individual mean RTs were not correlated ($r < 0.01$), we will report Analyses of Variance (ANOVA) for RT only, using GreenHouse–Geisser correction for *p*-levels; uncorrected degrees of freedom will be reported. Since mean RTs of illiterates were nearly entirely outside the range of those of normal adult readers (Lachmann and van Leeuwen, 2004, 2008b; van Leeuwen and Lachmann, 2004; Jincho et al., 2008), ANOVAs were run for each group separately.

For the illiterates group, a two-factors repeated measures ANOVA with Congruence (isolated, congruent surrounding, incongruent surrounding) and Material (letter vs. shape) as within-participant factors revealed a main effect for Congruence, $F_{(2,\ 62)} = 5.86$, $p < 0.01$, but not for Material ($F < 1$). No interaction was in evidence ($F < 1$). Congruent items (995 ms, SD = 475) were slower than isolated ones (967 ms, SD = 458), $F_{(1,\ 31)} = 5.67$, $p < 0.05$, and slower than incongruent ones (976 ms, SD = 429), $F_{(1,\ 31)} = 4.8$, $p < 0.05$. Isolated and incongruent items did not differ ($F < 1$).

The same ANOVA was run with the data from literate participants. Main effect were found for Congruence, $F_{(2,\ 50)} = 11.23$, $p < 0.01$, and Material, $F_{(1,\ 25)} = 32.66$, $p < 0.01$. Isolated items (443 ms, SD = 186) were faster than congruent (458 ms, SD = 183) and incongruent items (461 ms, SD = 187), which did not differ. For Materials, letters (433 ms, SD = 177) were responded to faster than shapes (476 ms, SD = 191). An interaction was found between Material and Congruence, $F_{(2,\ 50)} = 4.89$, $p < 0.01$, due to a negative congruence effect for letters and a positive congruence effect for shapes: Participants were faster with isolated letters (421 ms, SD = 178) than with incongruent letters, $F_{(1,\ 25)} = 7.99$, $p < 0.01$; Incongruent letters, in turn, were responded to faster than congruent ones (444 ms, SD = 175), $F_{(1,\ 25)} = 10.79$, $p < 0.01$. Also for shapes, a Congruence effect was observed, $F_{(2,\ 50)} = 5.69$, $p < 0.01$. Participants responded faster to isolated (467 ms, SD = 191) than to incongruent shapes (487 ms, SD = 195), $F_{(2,\ 50)} = 8.32$, $p < .001$ but not significantly faster than to congruent ones. Congruent shape (473 ms, SD = 186) were responded to faster than incongruent ones, $F_{(2,\ 50)} = 6.04$, $p < 0.05$. Results for both groups are displayed in **Figure 4**.

DISCUSSION

Amongst visual configurations, letters are special; practice and familiarity make that letters are processed much more efficiently than non-letter configurations of similar complexity (van Leeuwen and Lachmann, 2004; Burgund et al., 2006; Lachmann and van Leeuwen, 2007). In addition, letters have a specific functional relationship to phonemic representations, a relationship

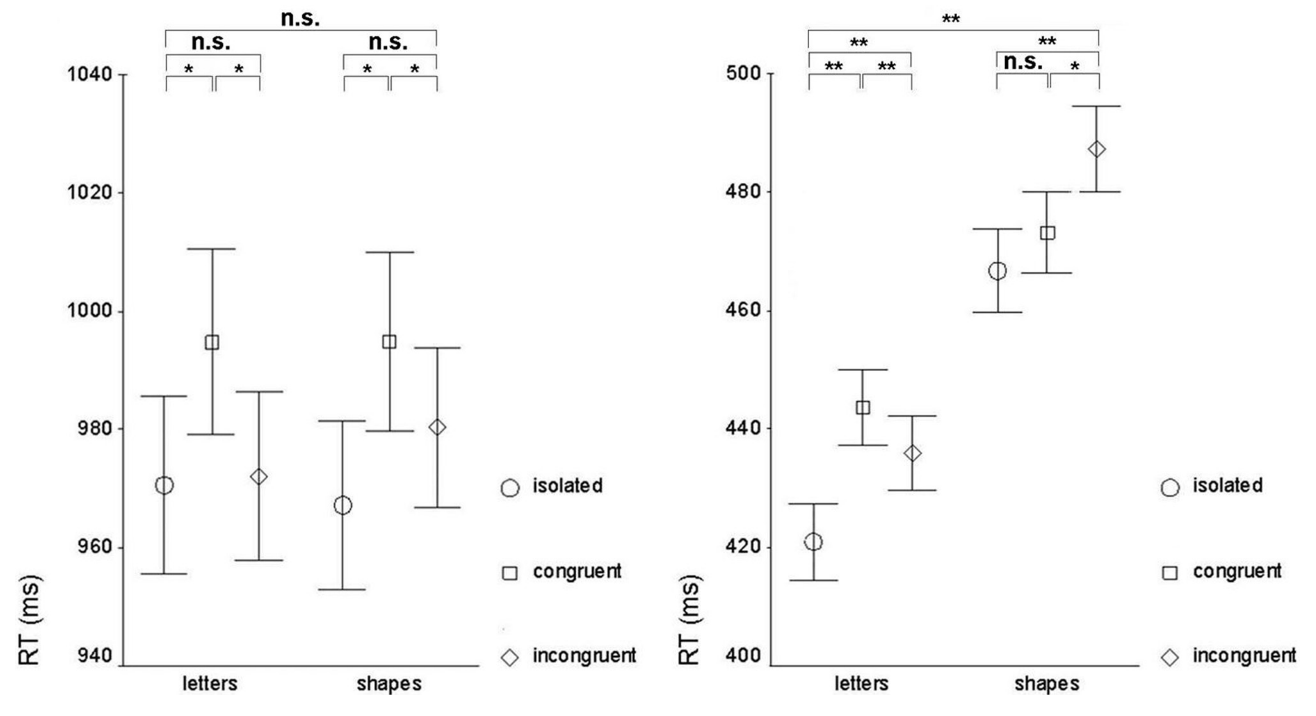


FIGURE 4 | Average reaction times (RTs) with error bars (5% confidence interval) for the experimental conditions for illiterate participants (left) and for literate control participants (* = significance at 5% level; ** = significance at 1% level; ns = no significant difference).

which is developed and automated while children learn to read (Frith, 1985). Recent studies have shown that this development leads to enhanced differentiation and responsiveness of the visual cortices to both orthographic and non-orthographic materials (Dehaene and Cohen, 2007; Dehaene et al., 2010a,b; Cantlon et al., 2011; Pegado et al., 2011). The differentiation gives a new role to established perceptual skills (Dehaene and Cohen, 2007; Lachmann and van Leeuwen, 2007; Blomert, 2011; Perea et al., 2011) which, after modification, need to be coordinated in order to guarantee fast and accurate reading (Lachmann, 2002). To automate this functional coordination, which in the end changes processing of both linguistic and non-linguistic stimuli (Lachmann and van Leeuwen, 2007; Dehaene et al., 2010b; Kolinsky et al., 2011), takes years of reading experience. We addressed the question: what has changed to our visual object perception, once this automatization process is completed?

We performed an experiment with illiterate adults, using the method by which in literates a differentiation between holistic non-letter processing and analytic letter processing was originally shown (van Leeuwen and Lachmann, 2004). This differentiation involves the way the letters are perceived in their immediate surroundings, not necessarily for how strongly their features are bound together at the within-object level (Lachmann and van Leeuwen, 2008b). The latter may depend on the Goodness of the object (Wagemans, 1993, 1999; van der Helm and Leeuwenberg, 1996), e.g., the symmetry of the letter “A.” On the other hand, observers tend to ignore symmetry in letters (Lachmann and van Leeuwen, 2007), suggesting that letters are also processed less holistically at this level. Neither does the differentiation in holistic processing at any of these levels have implications for the next higher level, which for letters would be that of morphemes or words. Our claim that letters are processed less holistically than non-letters, therefore, is not in conflict with the well-known word-superiority effect (Reicher, 1969). This effect applies at the level of groupings between letters. It could be argued that, in fact, recognition at this level might benefit from non-holistic processing at our current level (Freeman et al., 2003); word-level processes, for instance, will have difficulty matching individual letters of which the features have mistakenly been bound, based on pre-semantic information, to their surroundings. In sum, therefore, our claim of a distinction in holistic processing between letters and non-letters belongs exclusively to the level of visual integration between these objects and their immediate surroundings.

Even though both groups in the present experiment were of the same ethnicity they are likely to differ in more than their ability to read; such as in general intelligence (e.g., Neubauer and Fink, 2011), education and schooling (e.g., Tun and Lachman, 2008; Ventura et al., 2008), language skills (e.g., bilingualism; Martin-Rhee and Bialystok, 2008) and their familiarity with computers, which all may have influenced their performance. In particular, the latter factor may be responsible for the considerably higher reaction times over-all in illiterates compared to literates. Such differences are probably inevitable in these kinds of studies. However, the effects observed in our earlier studies in literates of different ethnicity and background contrast in the same, consistent manner with those in our illiterates. This may suggest that congruence effects are affected by literacy.

Illiterates not only were equally fast over all in processing letters and shapes, they also showed equally for both faster responses to targets presented in isolation versus in surrounding, as well as same congruence effects of the surroundings across letters and non-letters. For both letters and non-letters, incongruent surroundings led to faster responses than congruent ones.

In contrast, literates of the same ethnicity differentiated between letters and non-letters, just as groups of skilled readers of other ethnicities did (Lachmann and van Leeuwen, 2004; Jincho et al., 2008). They also showed a letter superiority effect that, not surprisingly, is absent in illiterates: letters are processed faster than non-letters and produce opposite congruence effects: positive congruence effects for non-letters, *negative* congruence effects for letters.

The negative congruence effect in illiterates, i.e., their preference for incongruent surroundings, implies that an analytic perceptual strategy prevailed. Developmental studies might have led us to expect that before reading is automated, at least in an age up from six years on, a holistic strategy to predominate (Schwarzer, 2002). Lachmann and van Leeuwen (2008a) compared adults with different groups of children: beginning normal readers from Grade 3 and 4, and age matched developmental dyslexics. Most of these children showed positive congruence effects for both letters and non-letters, indicating holistic preference. One reason could be that certain brain functions related to reading, especially auditory processing (Banai and Ahissar, 2006) are still developing at this age (Courchesne, 1978; Cheour et al., 2000; Shafer et al., 2000; Ceponiene et al., 2001; Parviainen et al., 2006, 2011; Bruder et al., 2010; Wetzel et al., 2011). This may keep them from using an analytic letter processing strategy that would enable rapid grapheme to phoneme encoding.

A subgroup of dyslexic children in Lachmann and van Leeuwen (2008a; see also Lachmann et al., 2010), with particular difficulties in reading non-words, however, show particularly strong negative congruence effects for letters. This suggests that at this stage in development the analytic strategy is at least present, even though, for object recognition and face recognition, the holistic strategy became already dominant (Schwarzer, 2002).

The results in dyslexics, therefore, are not inconsistent with those of illiterates: the negative congruence effects in illiterates means that analytic processing is not a reading-specific, secondary differentiation in perceptual organization that accompanies the process of learning to read. Rather, it is a generic and primary perceptual processing strategy, on a par with the holistic strategy (Schwarzer et al., 2010). Skilled reading recruits this general perceptual strategy for letter recognition, and uses it in a coordinated fashion along with other functions, including phonological, cognitive, motor, and attentional ones, in meeting the specific demands of reading. What is specific to skilled reading is not the automatization of a letter-specific perceptual strategy (Grainger et al., 2010; Pegado et al., 2011), but the (automated) coordination of various functional components specific to reading.

In this process, letter processing becomes habitually tied up with the analytic perceptual processing strategy. As a result, adult readers tend to no longer process simple non-letter objects analytically. This result is in accordance with the pervasiveness of congruency effects in visual object perception (Eriksen and Schultz, 1979; Pomerantz et al., 1989; Boenke et al., 2009). In incongruent conditions, observers fail to ignore irrelevant information, even if this would facilitate processing. This effect is usually considered a result of *attentional* interference of the irrelevant flanking or surrounding information; this remains a puzzle if we consider that, in principle, focused attention could have been applied to the target (Miles and Proctor, 2010). The present study suggests that this is because analytic processing has become associated with reading. Despite this, having learned to read does not render impossible the analytic processing of non-letter shapes. Evidence of analytic processing is not restricted to letters; negative congruence effects, although sparse, are found whenever active suppression of surrounding information is needed to distinguish a target (Briand, 1994; van Leeuwen and Bakker, 1995; Bavelier et al., 2000). Conversely, there is also evidence of holistic processing of letters if the task requires this (van Leeuwen and Lachmann, 2004). This implies that the association of analytic processing and letters is highly context-specific. The data suggest a gradual tendency for a specific, learned processing strategy to predominate for letters.

ACKNOWLEDGMENTS

Many thanks are due to Andrea Prölß and Andreas Schmitt (University of Kaiserslautern) for their assistance in the analyses of the paper-and-pencil-test and proof reading, to Rosana Tristão and Paul Nealen for helpful comments on earlier drafts of the manuscript, and to all colleagues and officials in India who helped us perform this study. Cees van Leeuwen was supported by an Odysseus grant from the Flemish Science Organization, FWO. Gunjan Khera was supported by a grant from the graduated studies program of the State of Rhineland-Palatinate (Germany).

REFERENCES

Banai, K., and Ahissar, M. (2006). Auditory processing deficits in dyslexia: task or stimulus related? *Cereb. Cortex* 16, 1718–1728.

Bavelier, D., Deruelle, C., and Proksch, J. (2000). Positive and negative compatibility effects. *Percept. Psychophys.* 62, 100–112.

Blomert, L. (2011). The neural signature of orthographic–phonological binding in successful and failing reading development. *Neuroimage* 57, 695–703.

Boenke, L., Ohl, F., Nikolaev, A., Lachmann, T., and van Leeuwen, C. (2009). Different time courses of Stroop and Garner effects in perception – an event-related potentials study. *Neuroimage* 45, 1272–1288.

Briand, K. A. (1994). Selective attention to global or local-structure of objects – alternative measures of nontarget processing. *Percept. Psychophys.* 55, 562–575.

Bruder, J., Leppanen, P. H. T., Bartling, J., Csepe, V., Demonet, J. F., and Schulte-Koerne, G. (2010). An investigation of prototypical and atypical within-category vowels and non-speech analogues on cortical auditory evoked related potentials (AERPs) in 9 year old children. *Int. J. Psychophysiol.* 79, 106–117.

Burgund, E. D., Guo, Y., and Aurbach, E. L. (2009). Priming for letters and pseudoletters in mid-fusiform cortex: examining letter selectivity and case invariance. *Exp. Brain Res.* 193, 591–601.

Burgund, E. D., Schlaggar, B. L., and Petersen, S. E. (2006). Development of letter-specific processing: the effect of reading ability. *Acta Psychol.* 122, 99–108.

Cantlon, J. F., Pinel, P., Dehaene, S., and Pelphrey, K. A. (2011). Cortical representations of symbols, objects, and faces are pruned back during early childhood. *Cereb. Cortex* 21, 191–199.

Ceponiene, R., Shestakova, A., Balian, P., Alku, P., Yiaguchi, K., and Näätänen, R. (2001). Childrens auditory event-related potentials index sound complexity and speechness. *Int. J. Neurosci.* 109, 245–260.

Cheour, M., Leppänen, P. H. T., and Kraus, N. (2000). Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clin. Neurophysiol.* 111, 4–16.

Courchesne, E. (1978). Neurophysiological correlates of cognitive development: changes in long-latency event-related potentials from childhood to adulthood. *Electroencephalogr. Clin. Neurophysiol.* 45, 468–482.

Dehaene, S., and Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron* 56, 384–398.

Dehaene, S., Nakamura, K., Jobert, A., Kuroki, C., Ogawa, S., and Cohen, L. (2010a). Why do children make mirror errors in reading? Neural correlates of mirror invariance in the visual word form area. *Neuroimage* 49, 1837–1848.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes, G., Jobert, A., Dehaene Lambertz, G., Kolinsky, R., Morais, J., and Cohen, L. (2010b). How learning to read changes the cortical networks for vision and language. *Science* 330, 1359–1364.

Eriksen, C. W., and Schultz, D. W. (1979). Information processing in visual search: a continuous flow conception and experimental results. *Percept. Psychophys.* 25, 249–263.

Freeman, E. D., Driver, J., Sagi, D., and Zhaoping, L. (2003). Top-down modulation of lateral interactions in early vision: does attention affect integration of the whole or just perception of the parts? *Curr. Biol.* 11, 985–989.

Frith, U. (1985). "Beneath the surface of developmental dyslexia," in *Surface Dyslexia; Neuropsychological and Cognitive Studies of Phonological Reading*, eds K. Patterson, J. Marshall, and M. Coltheart (London, UK: Erlbaum), 301–330.

Grainger, J., Tydgat, I., and Issele, J. (2010). Crowding affects letters and symbols differently. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 673–688.

Jincho, N., Lachmann, T., and van Leeuwen, C. (2008). Dissociating congruence effects in letters versus shapes: Kanji and kana. *Acta Psychol.* 129, 138–146.

Kolinsky, R., Verhaeghe, A., Fernandes, T., Mengarda, E. J., Grimm-Cabral, L., and Morais, J. (2011). *J. Exp. Psychol. Gen.* 140, 210–238.

Lachmann, T. (2002). "Reading disability as a deficit in functional coordination and information integration," in *Basic Functions of Language, Reading and Reading Disability*, eds E. Witruk, A. D. Friederici, and T. Lachmann (Boston, MA: Kluwer), 165–198.

Lachmann, T., and van Leeuwen, C. (2004). Negative congruence effects in letter and pseudo-letter recognition: the role of similarity and response conflict. *Cogn. Process.* 5, 239–248.

Lachmann, T., and van Leeuwen, C. (2007). Paradoxical enhancement of letter recognition in developmental dyslexia. *Dev. Neuropsychol.* 31, 61–77.

Lachmann, T., and van Leeuwen, C. (2008a). Different letter-processing strategies in diagnostic subgroups of developmental dyslexia. *Cogn. Neuropsychol.* 25, 730–744.

Lachmann, T., and van Leeuwen, C. (2008b). Differentiation of holistic processing in the time course of letter recognition. *Acta Psychol.* 129, 121–129.

Lachmann, T., Schumacher, B., and van Leeuwen, C. (2009). Controlled but independent: effects of mental rotation and developmental dyslexia in dual task settings. *Perception* 38, 1019–1034.

Lachmann, T., Steinbrink, C., Schumacher, B., and van Leeuwen, C. (2010). Different letter-processing strategies in diagnostic subgroups of developmental dyslexia occur also in a transparent orthography: reply to a commentary by Spinelli et al. *Cogn. Neuropsychol.* 26, 759–768.

Martin-Rhee, M. M., and Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Biling.-Lang. Cogn.* 11, 81–93.

Miles, J. D., and Proctor, R. W. (2010). Attention is required for acquisition but not expression of new response biases. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 1554–1560.

Neubauer, A., and Fink, A. (2011). Intelligence and neural efficiency: measures of brain activation versus measures of functional connectivity in the brain. *Intelligence* 37, 223–229.

Parviainen, T., Helenius, P., Poskiparta, E., Niemi, P., and Salmelin, R. (2006). Cortical sequence of word perception in beginning readers. *J. Neurosci.* 26, 6052–6061.

Parviainen, T., Helenius, P., Poskiparta, E., Niemi, P., and Salmelin, R. (2011). Speech perception in the child brain: cortical timing and its relevance to literacy acquisition. *Hum. Brain Mapp.* 32, 2193–2206.

Pegado, F., Nakamura, K., Cohen, L., and Dehaene, S. (2011). Breaking the symmetry: mirror discrimination for single letters but not for pictures in the Visual Word Form Area. *Neuroimage* 55, 742–774.

Perea, M., Moret-Tatay, C., and Panadero, V. (2011). Suppression of mirror generalization for reversible letters: evidence from masked priming. *J. Mem. Lang.* 65, 237–246.

Pomerantz, J. R., and Pristach, E. A. (1989). Emergent features, attention, and perceptual glue in visual form perception. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 635–649.

Pomerantz, J. R., Pristach, E. A., and Carson, C. E. (1989). "Attention and object perception" in *Object Perception: Structure and Process,* eds B. Shepp and S. Ballesteros (Hillsdale, NJ: Erlbaum), 53–89.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *J. Exp. Psychol.* 81, 275–280.

Schwarzer, G. (2002). Processing of facial and non-facial visual stimuli in 2-5-year-old children. *Infant Child Dev.* 11, 253–269.

Schwarzer, G., Kretzer, M., Wimmer, D., and Jovanovic, B. (2010). Holistic face processing among school children, younger and older adults. *Eur. J. Dev. Psychol.* 7, 511–528.

Serniclaes, W., Ventura, P., Morais, J., and Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition* 98, 35–44.

Shafer, V. L., Morr, M. L., Kreuzer, J. A., and Kurtzberg, D. (2000). Maturation of mismatch negativity in school-age children. *Ear Hear.* 21, 242–251.

Tun, P. A., and Lachman, M. E. (2008). Age differences in reaction time and attention in a national telephone sample of adults: education, sex, and task complexity matter. *Dev. Psychol.* 44, 1421–1429.

van der Helm, P. A., and Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: a nontransformational approach. *Psychol. Rev.* 103, 429–456.

van Leeuwen, C., and Bakker, L. (1995). Stroop can occur without Garner interference: strategic and mandatory influences in multidimensional stimuli. *Percept. Psychophys.* 57, 379–392.

van Leeuwen, C., and Lachmann, T. (2004). Negative and positive congruence effects in letters and shapes. *Percept. Psychophys.* 6, 908–925.

Ventura, P., Pattamadilok, C., Fernandes, T., Klein, O., Morais, J., and Kolinsky, R. (2008). Schooling in western culture promotes context-free processing. *J. Exp. Child Psychol.* 100, 79–88.

Wagemans, J. (1993). Skewed symmetry: a nonaccidental property used to perceive visual forms. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 364–380.

Wagemans, J. (1999). Toward a better approach to goodness: comments on van der Helm and Leeuwenberg (1996). *Psychol. Rev.* 106, 610–621.

Wetzel, N., Widmann, A., and Schröger, E. (2011). Processing of novel identifiablility and duration in children and adults. *Biol. Sci.* 86, 39–49.

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: 02 December 2011; accepted: 25 May 2012; published online: 12 June 2012.

Citation: Lachmann T, Khera G, Srinivasan N and van Leeuwen C (2012) Learning to read aligns visual analytical skills with grapheme-phoneme mapping: evidence from illiterates. Front. Evol. Neurosci. ***4****:8. doi: 10.3389/fnevo.2012.00008*

frontiers in
EVOLUTIONARY NEUROSCIENCE

REVIEW ARTICLE
published: 13 April 2012
doi: 10.3389/fnevo.2012.00005

A bird's eye view of human language evolution

Robert C. Berwick[1,2]*, Gabriël J. L. Beckers[3], Kazuo Okanoya[4] and Johan J. Bolhuis[5]

[1] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[2] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[3] Department of Behavioural Neurobiology, Max Planck Institute for Ornithology, Seewiesen, Germany
[4] Department of Cognitive and Behavioral Sciences, The University of Tokyo, Tokyo, Japan
[5] Behavioural Biology, Helmholtz Institute, University of Utrecht, Utrecht, The Netherlands

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

Reviewed by:
Christopher I. Petkov, Newcastle University, UK
W. Tecumseh Fitch, University of Vienna, Austria

***Correspondence:**
Robert C. Berwick, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 32D-728, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 32D-728, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
e-mail: berwick@csail.mit.edu

Comparative studies of linguistic faculties in animals pose an evolutionary paradox: language involves certain perceptual and motor abilities, but it is not clear that this serves as more than an input–output channel for the externalization of language proper. Strikingly, the capability for auditory–vocal learning is not shared with our closest relatives, the apes, but is present in such remotely related groups as songbirds and marine mammals. There is increasing evidence for behavioral, neural, and genetic similarities between speech acquisition and birdsong learning. At the same time, researchers have applied formal linguistic analysis to the vocalizations of both primates and songbirds. What have all these studies taught us about the evolution of language? Is the comparative study of an apparently species-specific trait like language feasible? We argue that comparative analysis remains an important method for the evolutionary reconstruction and causal analysis of the mechanisms underlying language. On the one hand, common descent has been important in the evolution of the brain, such that avian and mammalian brains may be largely homologous, particularly in the case of brain regions involved in auditory perception, vocalization, and auditory memory. On the other hand, there has been convergent evolution of the capacity for auditory–vocal learning, and possibly for structuring of external vocalizations, such that apes lack the abilities that are shared between songbirds and humans. However, significant limitations to this comparative analysis remain. While all birdsong may be classified in terms of a particularly simple kind of concatenation system, the regular languages, there is no compelling evidence to date that birdsong matches the characteristic syntactic complexity of human language, arising from the composition of smaller forms like words and phrases into larger ones.

Keywords: birdsong, brain evolution, phonological syntax, speech

INTRODUCTION: BIRDSONG AND HUMAN LANGUAGE PERSPECTIVES

Over 2000 years ago, Aristotle in his *Historia Animalium* (Aristotle, 1984, c. 350 BCE) had already noted many striking parallels between birdsong and human speech – in remarkably modern terminology, he observed that some songbirds, like children, acquire sophisticated, patterned vocalizations, "articulated voice," sometimes learned, and sometimes not: "second only to man, some species of birds utter articulate phonemes"; and "some of the small birds do not utter the same voice as their parents when they sing, if they are reared away from home and hear other birds singing. A nightingale has already been observed teaching its chick, suggesting that [birdsong] . . . is receptive to training" (*Hist. Anim.* 504a35–504b3; 536b, 14–20). In this passage, Aristotle uses the Greek word *dialektos* to refer to birdsong variation, paralleling the term he reserves for human speech, and anticipating even the most recent work on how the songs of isolated juvenile vocal learning finches might "drift" from that of their parents over successive generations (Feher et al., 2009). Given two millennia of research from neuroscience to genomics, our insights regarding the parallels between birdsong and human language have advanced since Aristotle's day. But how much have we learned? What can birdsong tell us today about the structure and evolution of human language?

In this article we consider this question from the perspective of modern linguistic theory, focusing on the connections between human language sound systems and syntax as compared to those of birdsong. We will maintain that while there are many striking parallels between speech and vocal production and learning in birds and humans, with both requiring similar, limited computational machinery, the same does not appear to hold when one compares language syntax and birdsong more generally. While there are many points at which birdsong and human syntax differ, summarized below in **Table 1** for reference, we highlight two here that seem especially prominent, returning to details and justification for this contrast in Section "Building Blocks for Human Language" below. First, human language syntax, but not birdsong, is organized into "chunks" – phrases – that are labeled by features of the elements from which the chunks are constructed (**Table 1**, row 7). For example, the word sequence *ate the starlings* has "verb-like" properties, inherited from the verb *ate*. In contrast, even though certain birdsong syllable sequences can be described as "chunks" (Suge and Okanoya, 2010), these do not have the properties of the

Table 1 | The major comparisons between birdsong syntactic structure and human syntactic structure.

	Birdsong	Human language syntax
Precedence-based dependencies (1st order Markov)	Yes	Yes, but in sound system only
Adjacency-based dependencies	Yes	Yes
Non-adjacent dependencies	In some cases	Yes
Unbounded non-adjacent dependencies	Not known	Yes
Describable by (restricted) finite-state transition network	Yes (*k*-reversible)	No
Grouping: elements combined into "chunks" (phrases)	Yes	Yes
Phrases "labeled" by element features	No	Yes (words)
Hierarchical phrases	Limited (in some species)	Yes, unlimited
Asymmetrical hierarchical phrases	No	Yes
Hierarchical self-embedding of phrases of the same type	No	Yes
Hierarchical embedding of phrases of different types	No	Yes
Phonologically null chunks	No	Yes
Displacement of phrases	No	Yes
Duality of phrase interpretation	No	Yes
Crossed-serial dependencies	No	Yes
Productive link to "concepts"	No	Yes

Most human language syntactic properties are not found in birdsong. The only exceptions relate to the properties of human language sound systems.

syllables out of which they are built; for example, the (hypothetical) chunk *warble-twitter* does not have the properties of either of the two syllables from which it is composed. Second, human language phrases are generally *asymmetrically hierarchical* (**Table 1**, row 9): the phrase *ate the starlings* is divisible into a small portion, the verb *ate*, and then a much larger portion, *the starlings*, which the larger portion might in turn contain further elaboration, as in *ate the starlings that sat on the wire*. Nothing like this syntactic complexity seems evident in birdsong.

Marler (1998) has advanced a very similar view in his contrast of "phonological syntax" or *phonocoding*, as opposed to "lexical syntax" or *lexicoding*. On Marler's account, songbirds exhibit only phonological syntax, that is, the stringing together of elements, sounds, according to some well-defined pattern, but without the meaning of the resulting sequence as a whole dependent on the meaning of its individual parts. In contrast, Marler argues that only human languages exhibit lexical syntax, that is, changes in meaning resulting from different combinations elements such as word parts, words, or phrases – *starling* means something different from *starlings*. Put another way, Marler notes that while both birdsong and human language are combinatorial, in the sense that they both assemble larger structures out of more basic parts, only human language is compositional, in the sense that the meaning of a word or sentence changes as we change its component parts.

In this article we have used Marler's distinction as the springboard for a more nuanced review of the differences between birdsong and human language, one that focuses on both details about computation and representation. From the standpoint of computation, the difference between birdsong and human language syntax has often been cast as a single, sharp formal difference in the computational machinery available to humans as opposed to birds (and other non-human species): all birdsongs can be described in terms what are technically called *regular languages* – languages that can be generated by a particularly simple kind of computational device called a finite-state automaton, while human languages are *non-regular* and fall outside this class, describable only by using more powerful computational devices. The distinction between regular and non-regular language is familiarly known as part of the *Chomsky hierarchy* (Chomsky, 1956), one formal way of partitioning the complexity of languages when viewed as a set of strings. However, we find that while the regular/non-regular distinction captures some of the differences between birdsong and human language, it is both too weak and too strong. As we describe in Section "Human Language and Birdsong: The Key Differences" below, this distinction is too weak, because it appears that all birdsong can be described by a far narrower class of regular languages, that turn out to be easily learned from examples, an important point if birdsong is to be learned from adult male tutors (Berwick et al., 2011a). But this distinction is also too strong, in the sense that several aspects of human language, such as the assignment of stress to words, or the way that prefixes or suffixes are assembled to form words, can be described by finite-state automata, while other aspects of human language seemingly go beyond the computational augmentations used to divide the regular from the non-regular languages (see, e.g., Huybregts, 1984).

In brief then, we find that from a computational perspective, the traditional Chomsky hierarchy does not draw the proper "bright line" separating human language from birdsong. (See Barton et al., 1987 for another view on the inadequacy of this hierarchy as a way to categorize human language.) Rather than impose an *a priori* classification on an inherently biological system such as language, drawn from the analysis of formal languages, the approach taken here turns the traditional classification on its head: we first attempt to characterize as best we can the minimally necessary computational components that empirically underpin language. Given this, we then characterize what class of sentences and structures this delimits. As **Table 1** indicates, human language must be analyzed at a finer grain than simply the regular/non-regular distinction.

Similarly, from a representational point of view, our characterization of how human language and birdsong differ in terms of asymmetrical, hierarchically arranged phrases does not fit neatly into any of the conventional categories fixed by the regular and non-regular languages, which do not typically address the question of what *structures* are assigned to particular strings. For example, as we shall see, the asymmetrical hierarchical structure associated with *ate the starlings* could just as readily be associated with a regular language as with a non-regular language. Finally, Marler's notion that it is "lexicoding" – words – that completely characterizes the division between human language and birdsong captures part, but not all, of the necessary distinctions. It does not account for the inherent asymmetry of human language structure, and falls short when it comes to describing human language structures that have no associated lexical meanings, such as the metrical or prosodic structure associated with human language.

Figure 1 sets out the gist of our account using a simple example sentence, where we have deliberately simplified linguistic details for expository purposes. **Figure 1A** displays the syntactic structure conventionally associated with the sentence *the rat chased the birds*. It exhibits two prominent properties. First, the representation is *hierarchical*. The sentence divides into two parts: on the left, the portion corresponding to *the rat*, ordinarily called a *Noun Phrase* (NP); and on the left, the portion corresponding to *chased the bird*, ordinarily called a *Verb Phrase* (VP). The VP itself then subdivides into two further parts, a verb *chased* on the left and a second Noun Phrase, *the birds*, on the right. Thus the first NP lies at one level *above* the second NP. This grouping of the verb and the second Noun Phrase together into a single unit, what linguists call a *phrase*, is not arbitrary. This analysis has been confirmed empirically for over a century, using established structuralist and generative linguistic techniques (see, e.g., Bloomfield, 1933; Chomsky, 1955; Jackendoff, 1977). For example, it is straightforward to show that the second Noun Phrase, *the birds*, conventionally called the Object, is bound together with the Verb as a single "chunk" or phrase, because the Verb plus its Object can be seen to be subject to syntactic rules that manipulate them single units, in the same sense that we identify particular combinations of atoms as specific molecules because they act identically in particular chemical reactions. Linguists have devised many standard tests to demonstrate the existence of such "chemical compounds" in language; we illustrate one of several here. Consider the example sentence (1a) below. Linguists note that the sequence *ate the birds* forms a single phrase, a verb phrase, because, as shown in (1b), one can remove the second occurrence of *ate the birds* in its entirety, substituting the word *did*, but retain the same meaning as in (1a), viz., that both the rat and the cat ate the birds. In contrast, if we delete any *part* of the "compound" Verb-plus-Object, and try to apply the same syntactic operation – the same "chemical reaction" – the result seems ill-formed, as evidenced by (1c):

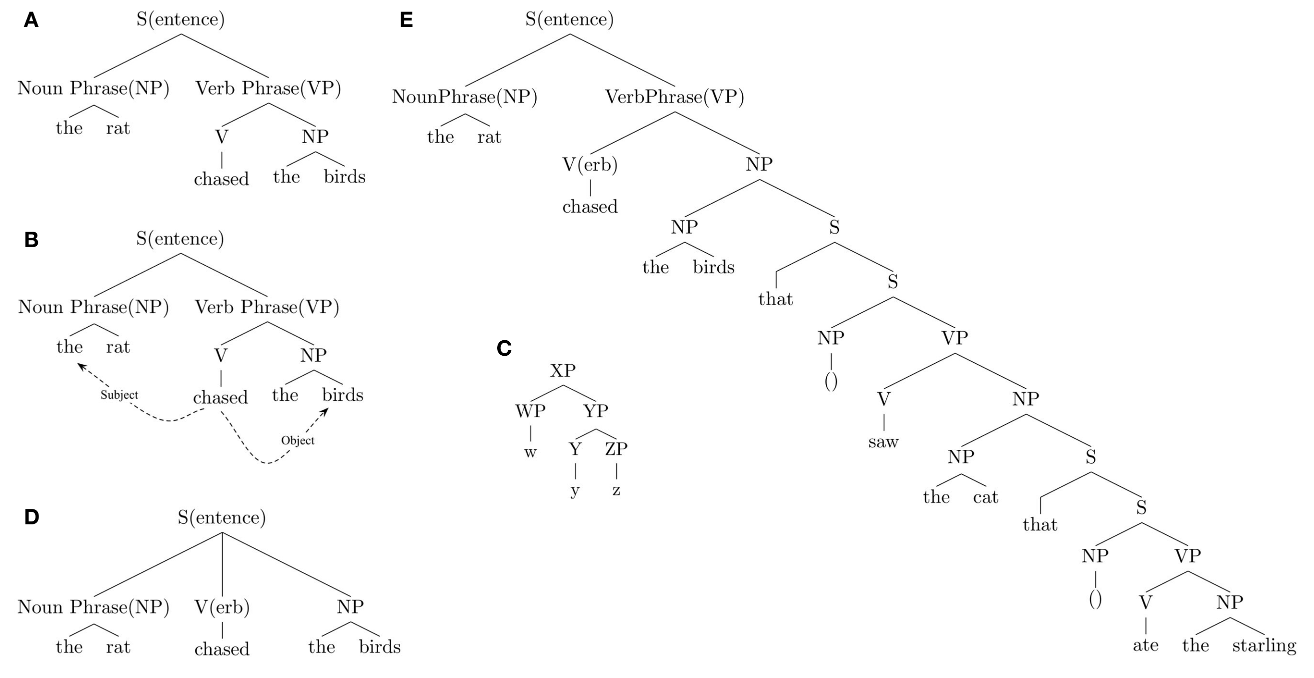


FIGURE 1 | The hierarchical and asymmetrical nature of sentence syntax. (A) The conventional syntactic structure associated with a simple English sentence. Note that the structure is asymmetrical, as highlighted in part **(C)** below. **(B)** The grammatical relationships "Subject" and "Object" are defined solely with respect to the hierarchical properties of the syntactic representation. **(C)** Abstract phrase structure representation for the sentence, highlighting the core asymmetry such structures, along with the "grounding" of a phrase of type *Y*, *YP*, on a particular word of type *Y*. Such structures comprise the core "molecular shapes" of syntax in language. **(D)** A symmetric syntactic structure associated with the same sentence, pointing out that there is no longer a distinction between "Subject" and "Object." **(E)** A more complex syntactic structure associated with a sentence that displays hierarchical, self-embedded containment relationships, with Sentences embedded within other Sentences, and Noun Phrases within Noun Phrases. Note that the basic asymmetric structure of **(C)** is replicated at each level.

(1a) *the rat ate the birds and the cat ate the birds*
(1b) *the rat ate the birds and the cats did too*
(1c) *?? the rat ate the birds and the cats did the birds too*
(1d) *the birds, the rat ate*

In this way, sensitivity to syntactic rules demonstrates that the Verb-plus-Object can be manipulated as if it were a single entity. Similarly, the sentence's Object, *the birds*, is itself a single unit, so it too can be manipulated as if it were a single syntactic "molecule": we can displace it to the front of a sentence, as in (1d). What about a hypothetical "compound" that would be formed by conjoining *the rat*, the so-called the Subject of a sentence, with the Verb, forming the unitary "molecule" *the rat ate*? Such a hypothetical unit is *never* observed to enter into distinguished syntactic operations – it is a "compound" that evidently does not participate in distinctive syntactic "chemical reactions." We may therefore conclude, along with the majority of linguists, that the "grouping" structure of words in English sentences like these may be portrayed in something like the form, Subject–Verb Phrase, where the Verb Phrase in turn is divided into a Verb plus its Object (if any). Because the Object itself forms a group, one is thereby licensed to represent the syntactic form of the entire word sequence as something like, (*the rat*) (*chased* (*the cat*)), where the Subject phrase is placed apart from the rest of the syntactic structure in the sentence, asymmetrically. It should be stressed that examples such as (1a–c) have also received confirmation from domains other than linguistic analysis, in this case, from psycholinguistic studies indicating that complete Verb Phrases, i.e., Verb–Object combinations, are "recycled" in human sentence processing, while there is no comparable evidence for this with respect to Subject–Verb combinations; see, e.g., Arregui et al. (2006), Mauner et al. (1995). For additional book-length treatment of the key role of asymmetric relations in language, see Kayne (1994), Moro (2000), Di Sciullo (2003).

In brief, language's syntactic structure is fundamentally asymmetric. **Figure 1A** illustrates this asymmetry graphically: the first NP, corresponding to *the rat*, lies off to the left side of the rest of the sentence, which is subsumed by the Verb Phrase. This fundamental asymmetry, cast in terms of a tree-structured representation as shown in **Figures 1A,B**, is central to how sentence structure drives sentence interpretation. The first NP directly dominated by the entire Sentence fixes what is the Subject, and this NP is typically, but not always, the "agent" of the action corresponding to the Verb. In contrast, the NP dominated by the VP and adjacent to the verb determines what is the Object, and this NP is typically the "affected object" of an action (Chomsky, 1965).

Importantly, such syntactic relationships do not depend on the temporal ordering of a sentence's words – the left-to-right way the words are orthographically transcribed, corresponding to their spoken (or manually signed) order. Rather, a considerable body of converging evidence, from linguistic, psycholinguistic, and more recently brain-imaging studies, has accumulated showing that this necessarily "linear" format is mapped to an internal representation that respects only hierarchical structure (see, e.g., Moro, 2008, 2011; for recent fMRI confirmation along these lines, see Pallier et al., 2011).

To take one additional example illustrating this point, consider the way that interrogative questions are formed in English, via the manipulation of the Subject and auxiliary verbs such as *is*. It was noted several decades ago by Chomsky (1965) that, given a declarative sentence such as, *the boy is sitting in the room*, the corresponding question form is given by, *is the boy sitting in the room*. Chomsky noted that the syntactic rule that forms such questions *cannot* be stated as, "displace the leftmost auxiliary verb to the front of the sentence." This is because, given a sentence where the Subject Noun Phrase contains another Sentence, such as *The boy who is sitting in the room is happy*, the corresponding question form works out as, *is the boy who is sitting in the room happy*; the corresponding question *cannot* be *is the boy sitting in the room is happy*. In other words, this syntactic rule does not pick out the *first* (as opposed to the *second* occurrence of *is*), but rather the *hierarchically most prominent* occurrence of *is*, the one that is part of the "main" sentence, *the boy is happy*.

More broadly, there is no known syntactic rule that operates on precisely the *third* element from the beginning of the sentence; that is, numerical predicates such as *third* or *fourth* are not part of the inventory of predicates in the human language syntactic system. Not only does this offer additional evidence on its own that human language syntactic structure is hierarchical, this hypothesis has been probed by psycholinguistic analysis. In a series of experiments, Musso et al. (2003) attempted to see whether there was a difference between the ability to acquire an artificial language rule that respected a numerical predicates, e.g., the formation of a question by placing a special word precisely *three* words from the start of a sentence, as opposed to a rule that respected more natural predicates for, e.g., question formation. The former type of rule they called a "counting rules." They found that such "counting rules" were indeed more difficult to acquire, being learned, if at all, as if they were "puzzles" as opposed to naturally occurring language patterns. In their later experiments, this finding was confirmed via brain-imaging: the "counting rules" activated distinctive brain regions that contrasted with those activated by "normal" linguistic rules. Unsurprisingly, the counting rules activated regions related to those also activated during non-linguistic puzzle solving. Similarly, Crain and Nakayama (1987) found that children acquired question formation rules that abided by hierarchical constraints, but never rules based on linear order.

Possibly, an even stronger position can be maintained. As far as can be determined, *all* syntactic relationships in human language syntax depend on the just the hierarchical properties of a sentence's structure, along with whether an item is simply adjacent to another item or not. Linear precedence is otherwise ignored. We present other evidence for this possibly surprising fact in Section "Human Language and Birdsong: The Key Differences" below. In contrast, in human speech (and in birdsong, as we suggest below), linear precedence *does* play a critical role; for example, in English, the past tense marker *ed* is placed at the end of a word, rather than the beginning, so that we say *chased* and not *edchase*.

The reason for decoupling human sound systems from human sentence syntax is that such key differences between spoken (or manually signed) language "output" and its internal representation bear critically on the comparison between birdsong and human language. While both birdsong and human language sound structures are linear, in the sense that left-to-right order, linear precedence, *does* matter, human language syntactic structure, drawing on hierarchical predicates, radically differs from birdsong. It is precisely here that one can pinpoint a "gap" between birdsong and human language. We return to this important point below, in Section "Birdsong Seems Analogous to Speech, Not Syntax."

Finally, as one more illustration of the hierarchical vs. linear contrast, note that the left-to-right order of the Subject, Verb, and Object in the example of **Figure 1A** is entirely particular to English. In other languages, for example, in Japanese, Bangla, and German, the Object would typically precede the verb. In this sense, the picture in **Figure 1A** might best be thought of as a mobile, with parts below the top, and at the two NP and hinge VP points, that can pivot around one another, interchanging, e.g., *the rat* with *chased the birds*. Such observed variation again underscores the fact that it is the hierarchical relationships that are central to syntax, rather than any left-to-right order.

If we now abstract away the details of the words and the names of the phrases, replacing them with labels like *XP* and *YP*, we arrive at **Figure 1C**, which highlights the basic asymmetry of human language syntactic structure. It displays a single asymmetric "molecule" structure virtually all current linguist theories posit at the heart of syntactic description. (This is true of even such otherwise divergent linguistic theories as Construction Grammar, Goldberg, 2006; Head-driven Phrase Structure Grammar, Sag et al., 2003; and modern generative grammar, Radford, 1997). Further note that the phrase *YP*, which in our example corresponds to a Verb Phrase, is partitioned into an element *Y*, corresponding to *chased*, plus another phrase, *ZP*, in our example, the Noun Phrase *the birds*. This reflects the important fact that a phrase of type *YP* is generally anchored on a word of the same sort *Y* in the way that a Verb Phrase is anchored on a Verb. We may contrast this kind of asymmetrical representation with a possible *symmetrical* structure assigned to the same sentence, depicted in **Figure 1D**, where the two Noun Phrases and the Verb are placed at one and the same level. While there is no difficulty with this representation in terms of separating out three components, NP, Verb, and NP, it is apparent that without additional information one cannot unambiguously determine which NP is the Subject, and which the Object, nor the demonstrable fact that the verb and the Object are more tightly bound together as if they were a single unit. In this sense, the symmetric representation is deficient. One could of course impose a linear ordering requirement on this triple of items to "solve" the problem of assigning the Subject and Object relations in this simple example, but this would not generalize to the full range of sentences, such as *the birds, the rat chased*. This is not to say that such structures are absent in language. For example, in conjunctions such as, *the starling ate fruit and insects*, the conjoined phrase *fruit and insects* can be reasonably construed as symmetrical – one can reverse the order to get *insects and fruit*, and obtain the same meaning. Nevertheless, asymmetrical structure remains the norm in human language. Indeed, there are evidently certain computational advantages to asymmetrical syntactic structure. For example, it has been observed since the work of Miller and Chomsky (1963), Chomsky (1963), Halle and Chomsky (1968), and Langendoen (1975), among others, that human language sentences are sometimes readjusted so as to render them asymmetric and easier to process. The classic example is the prosodic contour assigned to a sentence with several "embeddings" such as *this is the cat that bit the rat that chased the starlings*. The syntactic structure assigned to this sentence is deeply nested, as may be appreciated by its parenthetical syntactic representation, (*this* (*is* (*the cat* (*that chased* (*the rat* (*that* (*chased* (*the starlings*)))))))). However, interestingly, the sentence's prosodic contours do *not* follow the same syntactic format. Instead, there are strong intonational breaks that cut off after the asymmetrical first portion of each Subject is encountered, as may be indicated by vertical strokes: *the cat* | *that chased the rat* | *that chased the starlings* |. As emphasized by Langendoen (1975), it is as if the hierarchical structure has been "flattened," so rendering it easier to process by enabling a listener to process each chunk delimited by the vertical strokes before moving on to the next, rather than having to hold the entire Noun Phrase beginning with *the rat* in memory all at one time. Langendoen (1975) and Berwick and Weinberg (1985) suggest that this "chunking" is also partly semantic in character, in that the head word of each Noun Phrase (*cat*, *rat*, etc.) is seemingly interpreted semantically before "waiting" for the rest of the phrase (*that chased*. . .etc.) to be processed. In fact, Langendoen notes that this reflects part of a general processing strategy, what he calls "readjustment rules," that comprise some of externalization process referred to earlier. Further, there is an accumulating body of more recent results confirming the advantages of asymmetry in sentence processing; see, e.g., Fong and Di Sciullo (2005); and for a recent perspective from the perspective of neurolinguistics, confirming the basic asymmetry of language, see Friederici et al. (2011).

Though basic Subject/Object asymmetries have been confirmed by a substantial body of linguistic and psycholinguistic research, one line of experiment that has apparently not been attempted so far is in the area of artificial grammar learning. Here, the relevant questions have apparently yet to be pursued.

Why is this important for a comparison of human language and birdsong? It should also be evident that structures such as the one displayed in **Figure 1A**, accompanying the simplest of sentences, already carry us a long way from the domain of birdsong. As we describe in more detail below in Section "Human Language and Birdsong: The Key Differences," even the most complex birdsong does not use asymmetrical, hierarchical relations like that of "Subject" to fix its properties. Certain bird species such as nightingales apparently have quite complex songs which seem best described in terms of syllables linearly arranged into repeated "chunks," which are in turn arranged into song sections, then sections into packets, and finally packets into contexts (Todt and Hultsch, 1996). However, this kind of structure is neither asymmetrical nor built on combinations at one level that in turn constrain structure at one level above or below. We do not find that, say, the sequence of syllable chunks in a nightingale's song depend on the hierarchical structure of song sections or packets. This is in distinct contrast to the typical format of human syntactic structure illustrated above, where a verb that forms a Verb Phrase picks out a Noun Phrase one level *above* its structural level as the Subject. Rather, to reinforce the point made earlier, what (limited) hierarchical arrangements are found in birdsong seem fixed by a linear, left-to-right sequencing, unlike human syntax, but similar to human speech.

There is yet one more critical difference between birdsong and human language syntax, illustrated in **Figure 1E**. In human language, Sentences, Noun Phrases, and indeed phrases of any type, can be contained entirely within other Sentences, NPs, and phrases of other types *ad infinitum*. This was already illustrated by the example of question formation in an example such as, *the boy who is sitting in the room is happy*, where the phrase *the boy who is sitting in the room* is an example of a Noun Phrase *the boy*... that properly contains a sentence-like phrase, *who is sitting in the room*. Note that this kind of containment relation might be extended: we could have a sentence such as, *the boy who is sitting in the room that is on the top floor is happy*, where there are now *two* sentence-like objects contained within the Subject *the boy*. Since such sentence structures are asymmetrical, the basic asymmetrical "molecule" of **Figure 1B** is replicated at several different scales, in a self-similar, fractal-like way. Birdsong does not admit such extended self-nested structures, even in the nightingale: song chunks are not contained within other song chunks, or song packets within other song packets, or contexts within contexts. Moreover, there seems to be no evidence that distinguished structural containment relationships are manipulated in birdsong to yield distinct meanings in a way analogous to human language. In short, as **Figure 1A** indicates, such asymmetric containment relationships are basic to every sentence, the rule rather than the exception in human language.

In any case, the possibility of arbitrarily extended, labeled hierarchical structures in human language admits an open-ended number of internalized, distinct representations. In the remainder of this article, we will argue that birds seem to lack a comparable syntactic ability. This distinction remains even if one puts to one side the obvious fact that birds do not seem to have conceptual units like words, focusing purely on syntactic combinatorial abilities. While there is a single recent publication to the contrary suggesting that at least one species, Bengalese finches, might possess some facility at both learning and then perceiving open-ended hierarchical representations that fall into the class of so-called strictly context-free languages (Abe and Watanabe, 2011; see Section Birdsong Seems Analogous to Speech, Not Syntax below for a discussion of this terminology), the experimental design of this study is apparently flawed, as we discuss briefly below and as detailed in Beckers et al. (2012). This "gap" between human and avian syntactic abilities marks out a key difference between human language and birdsong, because an open-ended combinatorial syntax operating over atomic units (like words) has long been regarded as perhaps the hallmark of human language. Even though some have speculated otherwise (Petri and Scharff, 2011), there is no evidence that songbirds "name" and then re-use combinatorial units similar to *ate the birds* to arrive at an arbitrarily large number of combinatorial possibilities. **Table 1** in Section "Human Language and Birdsong: The Key Differences" brings together and summarizes all of these birdsong–human language comparisons.

In considering this summary comparison, we should emphasize that it would be a mistake to conclude that all birdsong–human differences result simply from the lack of words in birdsong, as we discuss further below. For example, even though birds lack words, there is nothing that logically blocks birdsong syntax from relying on syllable groupings or other features that could themselves be labeled by properties of their constitutive parts, which could then be assembled into more complex units in the same way that a Verb Phrase is labeled by the properties of the Verb it subsumes. Of course, this is a hypothetical example, since to the best of our knowledge no birdsong is in fact constructed in this manner. But examples like these illustrate that it is not the lack of words alone that blocks the possibility of more complex birdsong syntax. Rather, this gap is due to a fundamental deficiency in a very particular computational ability, namely, the lack of the combinatorial operation of the sort found in human language, as further described in Section "Human Language and Birdsong: The Key Differences."

Moreover, these distinctions between birdsong and human language do not entail that birdsong analysis can shed no light on human language. We conclude from our survey that birdsong currently serves best as our best animal model of language's "input–output" component, describing how language is externalized and to a certain extend acquired, along with associated auditory–vocal and motor learning behaviors, such as auditory–motor rehearsal and vocal learning by auditory feedback and reinforcement. While this certainly does not encompass full human sentence syntax, nevertheless such information seems quite valuable in focusing our understanding of how human language works, including important details as to how language is acquired and produced, in the same sense that an understanding of the input–output interfaces of a complex computer system constrains, at least in part, of the remainder of the system that lies beyond the input–output interfaces. For example, one currently fruitful line of research in child language acquisition has probed the nature of infants' abilities to acquire particular sound patterns and word boundaries in part via statistical regularities (e.g., Saffran et al., 1996; Shukla et al., 2011, among much other work). Since this acquisition process involves the "input–output" system of human language, it would seem that it is precisely here where the songbird animal model could prove most useful. Indeed, as emphasized by Yip (2006), there are many basic questions regarding the connection between human and animal sound systems that remain unanswered, such as the precise role of statistical and prosodic features in birdsong, and their possible connection to the human language case. In this way, a deeper understanding of birdsong might facilitate greater insight into the case of human language acquisition. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-months-old infants. Finally, it seems equally misguided to reject out of hand the value of the songbird model because the "externalization" of human language can involve modalities other than sound, as in manually signed languages. In fact, the contrary seems to be true, as noted by Berwick and Chomsky (2011), and by Petitto et al. (2004); Petitto (2005): the sensory–motor sequencing involved in the human sound system can be carried over in large measure to the domain of manually signed languages. For instance, just as the physical constraints of the word limits the human sound system to the expression of dual predicates in a strictly linear, as opposed to a simultaneous fashion, e.g., *the cat chased the birds and ate the birds*, signed languages apparently operate under many of the same constraints, notwithstanding the different physical channel that logically admits such simultaneous expression.

The remainder of this article is organized as follows. We first review the basic evolutionary and neurobiological background comparing songbirds and humans with respect to auditory–vocal learning and sensory-guided motor learning, with a focus on homologous brain regions and genetic regulatory systems. Next, we situate both birdsong and human language within a common "system flow diagram" that delineates three major components: an "external interface," a sensory–motor-driven, input–output system providing proper articulatory output and perceptual analysis; a combinatorial rule system generating asymmetrically structured hierarchical sentence forms, incorporating words; and an "internal interface," a system mapping between the hierarchical structures of sentence syntax and a conceptual–intentional system of meaning and reasoning, loosely called semantics. This flow diagram will enable us to see more clearly what distinguishes birdsong and human language. We follow this system breakdown with a more detailed comparison of birdsong and human language syntax. We will see that *all* the special properties of human language syntax discussed earlier, along with others outlined in Section "Human Language and Birdsong: The Key Differences," can be directly accounted for if one assumes the existence of a single, simple combinatorial operation, anchored on words or more precisely, word features. It is this operation that is apparently absent in birds, so far as we know. However, even though birds seemingly lack words, it does not follow that the combinatorial operator is necessarily absent in birds. For example, the combinatorial operator could still work on other elements, for example, syllables, in this way yielding the distinctive metrical patterning of sound melodies, rhythmic patterns, as suggested in the domain of human language by Halle and Idsardi (1995). However, for whatever reason, the operator does not appear to have been exploited this way in birds. It remains an open question as to whether a similar analysis would apply to birdsong metrical patterns; this then is a possibly crucial open research question where a non-human model might (speculatively) provide insight into its counterpart in human language. If birdsong were found to operate in a similar way to human metrical structure, this might provide precisely the required evolutionary "bridge," in the sense that the combinatorial operator was present in the common ancestor of both species, but full-fledged language required in addition words and their features, an ability present in the human lineage, but not in any bird species. It follows that it is precisely here that one might look for key evolutionary innovations that distinguish humans from birds, a topic we briefly address in our conclusion.

AN EVOLUTIONARY PERSPECTIVE: CONVERGENT MECHANISMS AND SHARED COMMON DESCENT BETWEEN BIRDS AND HUMANS

The most recent common ancestor of birds and mammals, originating from the clade *Amniotes*, lived about 300 million years ago (Laurin and Reisz, 1995). Thus, at least 600 million years of evolution separate humans from *Aves*, a considerable stretch of time even in evolutionary terms. Given this length of time, is not surprising that birds and humans might share traits both in virtue of common descent, as well as a result of independent, convergent evolution. For example, evidently both birds and mammals share keratin genes derived from their common ancestor, giving rise to both feathers and hair, while wings and flight were developed independently by birds and bats or flying squirrels. Unsurprisingly, some traits are also a complex blend resulting both from common descent and convergent evolution. For example, birds (and their ancestors) retain a superior color vision system that was apparently lost in mammals, and then only recently recovered by certain mammals, in part by multiple recent gene duplications or other tinkering of certain retinal photoreceptive opsin control regions that differ in important details even from primate to primate – one reason people, but not birds, can be colorblind (Dulai et al., 1999). Even more recently it has been shown that bats possess "superfast" laryngeal muscles for echolocation that can work at frequencies greater than 100 Hz; such muscles are also found in songbirds (Elemans et al., 2011). Note that while such laryngeal muscles are apparently not found in humans, there is other evidence for adaptations to speech; see Fitch (2010) for a comprehensive review. Such complexity of evolutionary patterning is worth bearing in mind when discussing the comparative evolution of sophisticated behavioral traits like birdsong and human language.

A complex interplay between convergent evolution and common descent even arises within the class *Aves* itself. From the most recent genomic evidence (Suh et al., 2011) it has been proposed that the capacity for vocal learning in passerine (oscine) birds such as the zebra finch and the non-passerine vocal learning birds such as parrots is more likely to have evolved in a common Psittacopasseran ancestor as a unique evolutionary event, leading to shared genetic/neural components enabling vocal learning, such as an anterior–medial vocal pathway as delineated by standard genome expressions studies (e.g., transcription factor expression studies; Jarvis and Mello, 2000; Jarvis et al., 2000). While this phylogenetic analysis remains controversial, on this account, hummingbirds developed their vocal learning abilities separately, as a result of convergent evolution. A similar comparative evolutionary analysis is not possible for humans, since no extant primates exhibit human vocal learning abilities. Consequently, absent evidence to the contrary, for the present it seems more secure to assume that, much like hummingbirds, vocal learning in humans is a convergent evolutionary trait, with clear specializations for both auditory/motor sequencing and vocal learning and imitation. Earlier hypotheses that certain components of the vocal tract have been uniquely adapted for human speech, such as a descended larynx, now seem questionable (Fitch, 2005). More recently it has been argued that the convergent specializations for human vocalization and speech seem to lie at a deeper neural level and involve, among other components, a capacity for vocal imitation (Fitch, 2005). The recent findings regarding the role of a specific regulatory protein, Foxp2, in motor sequencing, addressed below, reinforce this view.

Turning to the interplay between common descent and convergent evolution, over the past decade many studies have confirmed that songbirds and humans possess homologous brain regions for auditory–vocal and motor-driven learning (Jarvis et al., 2005). There are several neural and genetic parallels between birdsong and speech (Bolhuis et al., 2010). The songbird brain has two interconnected neural networks, involved in song production, perception, and learning, as depicted in **Figure 2** (Bolhuis and Eda-Fujiwara, 2003, 2010; Bolhuis and Gahr, 2006; Jarvis, 2007; Bolhuis et al., 2010). First, secondary auditory regions, including the caudomedial nidopallium (NCM) and caudomedial mesopallium (CMM; **Figure 2A**), are involved in song perception and are important for the recognition of tutor song (Bolhuis and Eda-Fujiwara, 2003, 2010; Moorman et al., 2011). Second, the "song

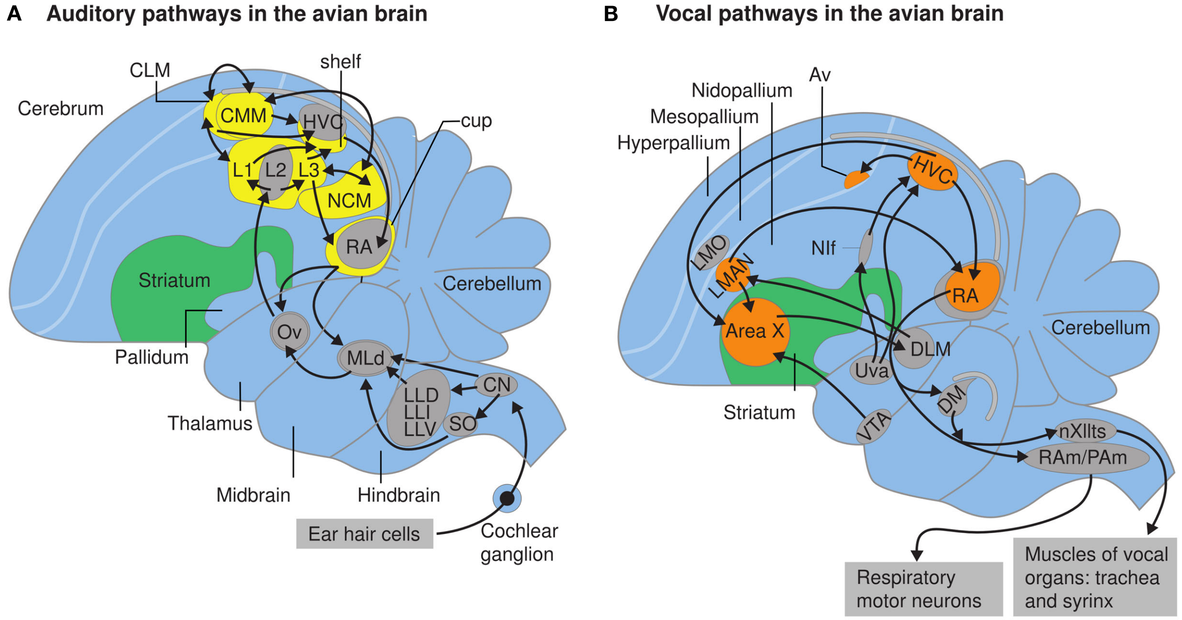


FIGURE 2 | Schematic diagrams of composite views of parasagittal sections of the songbird brain. (A) Diagram of a songbird brain giving approximate positions of nuclei and brain regions involved in auditory perception and memory. Yellow areas represent brain regions that show increased neuronal activation when the bird hears song. **(B)** Diagram of a songbird brain giving approximate positions of nuclei and brain regions involved in vocal production and sensorimotor learning. The orange nuclei in the song system show increased neuronal activation when the bird is singing (see text for details). Abbreviations: Cb, cerebellum; CLM, caudal lateral mesopallium; CMM, caudal medial mesopallium; DLM, nucleus dorsolateralis anterior, pars medialis; HP, hippocampus; HVC, acronym used as a proper name; L1, L2, L3, subdivisions of Field L; LaM, lamina mesopallialis; LMAN, lateral magnocellular nucleus of the anterior nidopallium; mMAN, medial magnocellular nucleus of the anterior nidopallium; NCM, caudal medial nidopallium; nXIIts, tracheosyringeal portion of the nucleus hypoglossus; RA, robust nucleus of the arcopallium; V, ventricle. Modified and reproduced, with permission, from Bolhuis et al. (2010) and Moorman et al. (2011), copyright 2010 Nature Publishing Group. All rights reserved.

system" is involved in song production and certain aspects of song learning (**Figure 2B**). The song system is subdivided into two major pathways, the song motor pathway (SMP; Mooney, 2009) and the anterior forebrain pathway (AFP; Brainard and Doupe, 2000; Doupe et al., 2005). The SMP is a posterior motor pathway connecting the HVC (acronym used as a proper name), the robust nucleus of the arcopallium (RA) and the tracheosyringeal portion of the nucleus hypoglossus (nXIIts), and is important for song production. The AFP is an anterior cortical–basal ganglia–thalamic loop that originates in HVC and passes through Area X, the thalamic nucleus dorsolateralis anterior, pars medialis (DLM) and the lateral magnocellular nucleus of the anterior nidopallium (LMAN), and eventually connects with the motor pathway at the nucleus RA. The AFP is essential for sensorimotor learning and adult song plasticity (Brainard and Doupe, 2002; Mooney, 2009).

In humans, conventionally the neural substrate of motor representations of speech is thought to involve Broca's area in the inferior frontal cortex, while perception and memory of speech is considered to involve Wernicke's area and surrounding regions in the superior temporal cortex. Although there are considerable differences between avian and mammalian brains, there are many analogies and homologies that have recently prompted a complete revision of the nomenclature of the avian brain (Jarvis et al., 2005). Similarities in connectivity and function would suggest at least analogies between the human neocortex and the avian pallium (including the hyperpallium, mesopallium, nidopallium, and arcopallium; see **Figure 2A** (Bolhuis and Gahr, 2006; Bolhuis et al., 2010). Specifically, Bolhuis and Gahr (2006) have suggested that the NCM and CMM regions in the songbird brain may be analogous with the mammalian auditory association cortex. In addition, Doupe et al. (2005) have argued that the AFP loop in the song system (**Figure 2B**) bears strong similarities in connectivity, neurochemistry and neuron types to the mammalian basal ganglia, while both LMAN and HVC have been suggested to be functionally similar to Broca's area (see Bolhuis et al., 2010 for further discussion).

In addition to these neuroanatomical parallels, there is increasing evidence for a similar neural dissociation between auditory recognition and vocal production regions in the brains of songbirds and humans (Gobes and Bolhuis, 2007; Bolhuis et al., 2010). Regions in the songbird caudomedial pallium (including the NCM) contain the neural representation of tutor song memory that is formed in juvenile males (Bolhuis and Gahr, 2006), whereas nuclei in the song system are required for sensorimotor learning and song production (Brainard and Doupe, 2000). Lesions to the NCM of adult zebra finch males impaired recognition of the tutor song, but did not affect song production, while lesions to the HVC in songbirds disrupted song production, but lesions to the nidopallium and mesopallium did not (Gobes and Bolhuis, 2007; Bolhuis et al., 2010). These and other findings suggest that in songbirds there is a neural dissociation between song recognition and song production that is already apparent in juveniles (Gobes and Bolhuis, 2007; Gobes et al., 2010). In human speech there is a comparable dissociation between brain regions involved in auditory perception and memory on the one hand, and vocal production on the other. Human newborns show increased neural activity in the superior temporal lobe, but not in the inferior frontal cortex, in response to human speech (Imada et al., 2006), while 3- to 12-month-old infants showed activation in both Wernicke's and Broca's areas in response to hearing speech (Dehaene-Lambertz et al., 2006; Imada et al., 2006). Taken together, these studies suggest that Wernicke's area is (part of) the neural substrate for speech perception in neonates and that Broca's area becomes active at a later stage, when infants start babbling; see Bolhuis et al. (2010), Brauer et al. (2011).

It is not yet completely clear whether these neural structures and information processing pathways are the result of shared ancestry, and so represent instances of homology, as opposed to convergent evolution and so analogy. Much remains to be understood about the detailed genetic, developmental, and neural underpinnings of vocal learning and language in both species. One key genetic parallel between birdsong and speech involves FOXP2, the first gene specifically implicated in speech and language (Fisher and Scharff, 2009). This is an ancient gene that codes for the transcription factor FoxP2, a protein that regulates DNA expression. Mutations in this gene in a large three-generation family and in some unrelated individuals were found to correlate with a speech disorder (Fisher et al., 1998). FOXP2 sequences are highly conserved between birds and mammals, and FOXP2 mRNA is expressed in song nuclei in the three known orders of song learning birds. FOXP2 is developmentally and seasonally regulated in songbirds and intact FOXP2 levels are required for normal song learning (Fisher and Scharff, 2009). As noted by Scharff and Petri (2011), this system may be part of a "molecular toolkit that is essential for sensory-guided motor learning" in the relevant brain regions of both songbirds and humans. Depressed vocal learning in songbirds that has been attributed to FoxP2's role in regulating other genes involved guiding neuronal development (Haesler et al., 2004; Vernes et al., 2011). In this sense, FoxP2 serves as an example of "deep homology" – a shared trait involved as part of both human speech and songbird vocal learning (Bolker and Raff, 1996; Shubin et al., 1997; Fitch, 2011; Scharff and Petri, 2011). However, the scope of this homology must be considered with some care. Since both vocal learning and non-vocal learning birds possess identical FoxP2 genes (Webb and Zhang, 2005), and the birds' FoxP2 genes are distinct from those of humans, differences in this gene alone cannot be what accounts for the vocal learning/non-learning distinction in birdsong. Rather, this difference seems to reflect differential gene expression as part of some larger overall gene network, as Haesler et al. (2004, p. 3174) note, "*FoxP2* has a characteristic expression pattern in a brain structure uniquely associated with learned vocal communication, Area X in songbirds." From this point of view, FoxP2 comprises one of probably many necessary ingredients in a complex recipe for vocal learning and production, rather than a single "master gene" that sits at the top of a regulatory cascade as in the case of the well-known regulatory *Pax*-6 *eyeless* homeobox gene (Halder et al., 1995).

BUILDING BLOCKS FOR HUMAN LANGUAGE

To better frame a comparison between birdsong and human language, it is helpful to partition language's fundamental relationship between sound and meaning into three distinct components: (1) an input–output system encompassing how language is produced, either acoustically, by vocal production, or manually, by signed gestures, as well as how language is perceived, by the auditory or visual system; (2) an internal rule system generating legitimate organism-internal structured representations, including, but not limited to, the kinds of structures depicted in **Figures 1A,E**, and (3) a system interfacing to cognitive processes such as meaning and inference, often glossed as "semantics." The first component includes representations such as the placement of stress that are not strictly sensory–motor in nature. In current linguistics, this component includes both acoustic phonetics and phonology. The second, rule-governed component feeds the other two, both the input–output interface as well as the semantic interface. This division is by no means universally accepted. For example, some linguistic accounts reduce or eliminate the role of a distinctive syntactic component, instead assuming a more direct relationship between sound and meaning (e.g., Culicover and Jackendoff, 2005; Goldberg, 2006; Jackendoff, 2010).

For example, Jackendoff (2010) argues that both components (1) and (3) have additional, separate interfaces to the mental repository of information about words, the lexicon, bypassing the syntactic component (2). Such additional links are quite plausible, because words – lexical items – have both phonological and semantic aspects, their particular sounds and meanings. In Jackendoff's view, such a division lends itself to a more natural evolutionary account where sounds and meanings might similarly be directly connected, without the intervention of syntax, this possibly serving as a kind of "protolanguage" stage. On the other hand, this position requires that there be an independent generative component for semantic representation, one that, according to Jackendoff, antedated human language syntax. At the moment, there seems to be little hard evolutionary evidence to distinguish between such alternatives, and in any case, the three-way division suffices for the bird–human comparison. This three-way dissection does factor apart the distinct knowledge types and representations generally recognized as central to language, in one way that enables a fruitful comparison.

BIRDSONG SEEMS ANALOGOUS TO SPEECH, NOT SYNTAX

Referring then to these three components, it is important to respect both the similarities and the differences between human speech and the totality of human language on the one hand, and birdsong on the other, which can and have led to some common misunderstandings. While speech is one prominent component of human language, it is neither necessary (as manually signed languages illustrate) nor sufficient. Rather, human speech, or more precisely, the sequenced motor commands involving a small number of vocal articulators such as the tongue, lips, velum, and larynx, comprises the end product of more sophisticated cognitive computations that engage at least two additional components: first, an internal combinatorial syntax; and second, a mental representation of both individual words and their meanings as determined by a particular syntactic combinations.

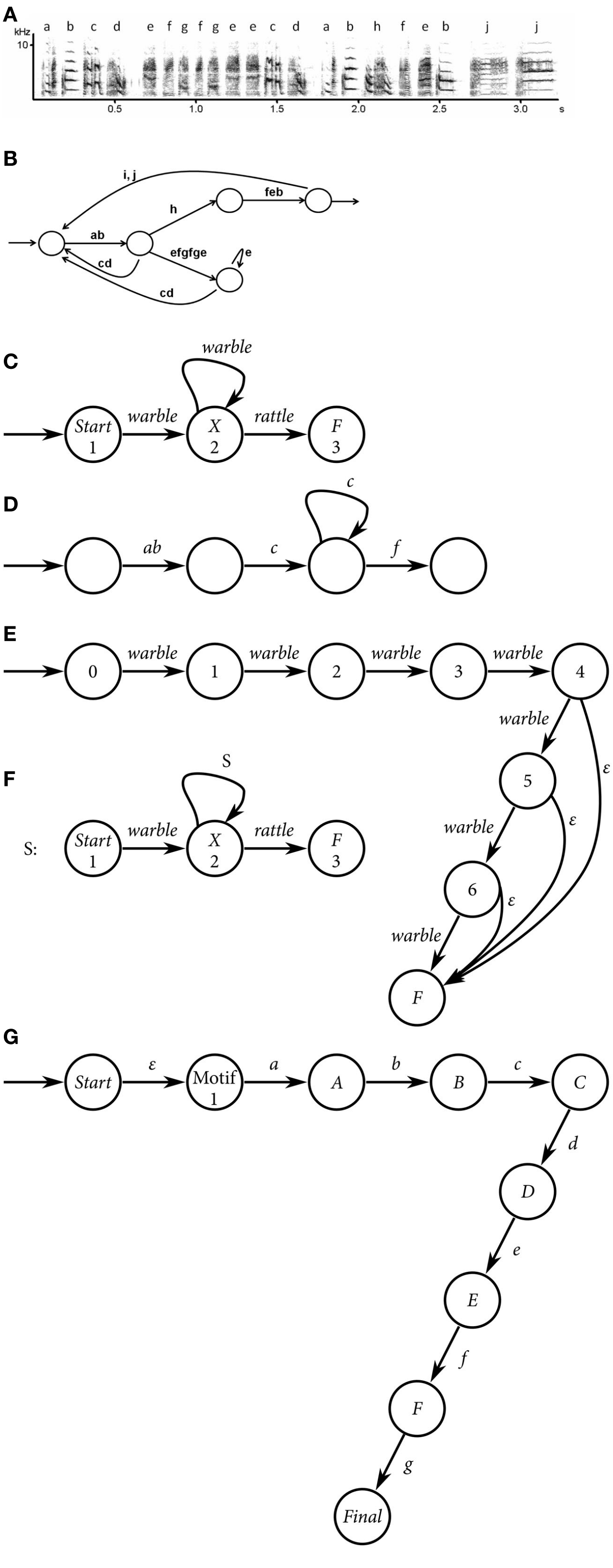


FIGURE 3 | Continued

FIGURE 3 | Continued
(A) Sonogram of an adult male Bengalese finch. *X*-axis is in seconds, *Y*-axis in kilohertz. Song syllables are demarcated by alphabet letters. **(B)** Finite-state transition network corresponding to the song syllable sequence in **(A)**. The network begins at the extreme left. Open circles correspond to states in the network, with transitions on arcs labeled with the syllables corresponding to those identified from the sonogram in **(A)**. Note that loops in the network can go back to previous states. **(C)** A finite-state transition network that generates syllable sequences containing at least one or more *warbles*, ending with a *rattle*. **(D)** A finite-state transition network encoding an "unbounded" dependency, in the sense that a syllable sequence beginning with *ab* must always end with exactly a single *f*. Note that syllable sequences may be arbitrarily long, due to the loop labeled with a *c* from one state back to itself. Thus, even though the *ab* and *f* may be arbitrarily far apart, a finite-state network can still determine whether this constraint holds. **(E)** A finite-state transition network that "counts" any number of *warbles* between four through seven, inclusively. The transitions labeled with ε denote so-called "epsilon transitions" where an output syllable is not produced when moving between states. **(F)** A recursive transition network labeled *S* that uses *S* itself on the transition looping from state 2 back to state 2 as a subroutine to generate an indefinitely large number of properly nested *warble-rattle* pairs. States are numbered for convenience. **(G)** A finite-state transition network that describes a hypothetical zebra finch song motif, as represented by a sequence of seven syllables, *a* through *g*. Note that if there are no nested dependencies, then the state *Motif* could be reached from any other state as part of a larger network describing the overall song.

In order to meet the demands of real-time speech/signed language production, in some way the human language system must map structured syntactic word combinations onto a sequence of motor commands, feeding a sensory–motor articulatory/gestural system for vocal or signed output, "flattening" the structure onto the output channel so that vocal output is sequentially ordered; see Stevens (2000). Conversely, the human processor recovers hierarchical structures from a time-ordered sound sequence. We might call this output projection *externalization*. It is typically here that *linear precedence* relationships hold among word elements in regards to their output as articulatory sequences, as was noted in the Introduction. Importantly, the detailed study of human sound systems has established that only linear precedence relations are required for the description of such systems; see Heinz and Idsardi (2011) and Wohlgemuth et al. (2010) for further discussion. To consider another simple language example here, the plural marker for *apple*, the so-called *z* morpheme in English, is placed at the end of *apple*, rather than the front, yielding *apples* (pronounced *applez*), rather than *zapple*. Conversely, if one regards the perception of language as mapping the time stream of acoustic signals into an internal representation, one must invert this process, recovering the hierarchical structures associated with sentences from the "flattened" signal.

From this standpoint, it is misleading to equate birdsong vocal production with the totality of human language. As we will now argue in some detail, birdsong seems more comparable to human language sound systems, not human language syntax. As we will argue, both human and bird sound systems are describable solely in terms of a network of what basic sound elements can come before or after one another – either syllable chunks in the case of birdsong, or so-called phonemes in the case of human language.

We will formalize this intuition below as the notion of a *finite-state transition network*.

What does this difference amount to descriptively? For birds, songs may consist of individual notes arranged in order as syllable sequences, where a syllable is defined, contrary to its usual meaning in linguistic theory, as a sound preceded and followed entirely by silence. Birdsong syllables, in turn, may be organized into recognizable sequences of so-called "motifs," and motifs into complete song "bouts." In some cases, the description seems to require more complexity than this, a matter considered in detail in what follows. Depending on the songbird species, the motif arrangements and ordering vary greatly, with the transitions between motifs probabilistic. For example, starling song bouts may be composed of many distinctive syllabic motifs lasting 0.5–1.5 s, up to a total length of 1 min (Gentner and Hulse, 1998), while nightingale songs consist of fixed 4-s note sequences, but arranged into a number of distinctive "chunks" with up to 200 distinctive song types. Supporting this view, Gentner and Hulse (1998) found that a first-order Markov model is sufficiently complex to describe possible starling song sequences.

Figure 3A displays a representative sonogram of a Bengalese finch song, with distinguishable syllables labeled as *a*, *b*, *c*, and so forth. By assembling a large sample of this bird's songs, one can extract a corresponding state diagram description as exhibited by **Figure 3B**. This picture consists of a finite, ordered sequence of states, the open circles, with transitions between the states labeled either by certain single syllable sequences, such as *h*, or multiple syllable units such as *ab* or *efgfge*. There are also loops that can carry one back to previous states, such as the syllables *i* or *j* that return to the leftmost open-circle state. By tracing out a syllable sequence starting from the entering arrow at the leftmost circle in the transition network, through to the exit arrow on the right, the network spells out or *generate* the entire set of legitimate syllable sequences for this bird's song repertoire, e.g., *ab efgffge cd ab h feb*. Note that even though there are only a finite number of states in this network, because of loops between some states, there can be a countably infinite number of valid possible paths from the starting arrow to the end state. To capture a bird's behavioral repertoire, typically these transitions are defined probabilistically, so that a transition between states occurs only with some positive probability corresponding to the likelihood of observing such a transition in the behaviorally observed data (Kakishita et al., 2009).

Such descriptions are conventionally called *finite-state transition networks* (Kleene, 1956); see **Figure 3C**. We now situate these within the standard framework of formal language theory (Hopcroft and Ullman, 1979). Here, a *language* is defined as any set of strings, equivalently, sentences, defined over a (fixed) alphabet, where the alphabet consists for example of the distinct syllables in a birdsong, or the distinct words in a human language. So for example, we might describe a particular birdsong "language" as consisting of "sentences" with any number of *warble* syllables *w* followed by an ending coda syllable, *rattle*, *r*. Such a birdsong language would contain an infinite number of "sentences," or songs, *wr*, *wwr*, *wwwr*, and so forth. Formally, languages are said to be *generated* by *transition networks*, where a finite-state transition network is a directed, labeled graph, consisting of a (finite) set of states, the nodes in the graph, connected by directed, labeled arcs, the edges of the graph. The notion of *generation* means that one can traverse the graph, beginning at a single designated *Start* state (denoted by a single incoming, unlabeled arrow in **Figures 3B,C**), and ultimately arriving at one or more designated *final* states. Generated sentences correspond to the sequence of labels on the edges arising during graph traversal. The set of all such possible label sequences from the *Start* state to a *final* state constitutes the *language generated* by the transition network. For present purposes, we need consider only two distinct types of networks: first, the *finite-state transition networks*; and second, a more powerful type of network, the *recursive transition networks* (Woods, 1970). (There is an equivalent approach that can be framed in terms of rule systems called *grammars*, either *regular* grammars, corresponding to the finite-state transition networks; or *context-free* grammars, corresponding to the recursive transition networks.)

We first consider finite-state transition networks and the languages they can generate. Finite-state transition networks can enforce the constraint that all syllable strings begin and end with one *warble*, or, following our earlier example, that a song contains any positive number of warbles, and end with a special final syllable *rattle*. The finite-transition network displayed in **Figure 3C** displays a finite-transition network obeying this second constraint. Let us see how. Generation begins at the *Start* state, also numbered 1. If we make a transition along the directed edge labeled *warble* to the state *X* (numbered 2), the system generates the first syllable in a possible output string, a *warble*. From state *X* there are two possible directed edges in the graph: one that leads back to state *X*, labeled with *warble*, and the other leading to the (single) distinguished final state *F* (numbered 3), labeled with *rattle*. If we take the transition labeled with *warble* back to state *X*, the generated sequence includes a second warble, and we can clearly continue in this way to output any number of *warbles* by traversing this loop any number of times. As soon as the system makes the transition from state *X* to state *F*, the syllable sequence ends with *rattle*, as desired. Note that the language so generated contains an infinite number of legitimate syllable strings, even though the network itself is entirely finite. It is in this sense that a finitely represented object can compute an extensionally infinite set of possible sentences.

More generally, the set of all finite-state transition networks generate the (syllable) stringsets called the *regular languages*, equivalently, stringsets defined by *regular expressions* (McNaughton and Yamada, 1960). Dependencies encoded by the regular languages can appear quite complex, including dependencies between items that are arbitrarily far apart from each other, what are sometimes called "unbounded dependencies." For example, the set of strings that begin with the syllable chunk *ab*, and then are followed by any positive number of *c* syllables, ending with an *f* that matches up with the beginning *ab*, can be described with via the regular expression $abc^{+}f$, where the $^{+}$ symbol denotes "1 or more occurrences." This language thus expresses an "agreement" constraint between the first and last syllables of any legitimate syllable sequence, even though there can be an indefinite number of *c*'s between the leading *ab* and the final *f*. Such "unbounded dependency" sequences can be generated by a very simple finite-state transition network with just four states, as displayed in **Figure 3D**. Petersson et al. (2012) are thus correct to point out that "the phenomenon of non-adjacent dependencies… can not simply be reduced to a choice between regular [i.e., finite-state transition network] or non-regular grammars [i.e., recursive transition networks].” However, as described in the introduction and as we pursue in more detail below, the phenomenon of *containment* of one type of phrase within a phrase of another type, when carefully articulated, *can* adjudicate between these two types of rule systems.

It appears that finite-state transition networks suffice to describe all birdsong. Indeed, it remains unclear whether birdsong even contains unbounded dependencies of the sort described in the previous paragraph, if we follow the results of Gentner and Hulse (1998) and others that first-order Markov processes, a more restricted network system, suffices to describe birdsong. (For a more recent confirmation of this claim, see Katahira et al., 2011.)

There are some apparent exceptions that merit additional discussion. Researchers have observed that the songs of certain bird species, such as chaffinches, consist of sections that must contain a particular *number* of iterated syllables of a certain sort, e.g., between 4 and 11 *warbles* (Riebel and Slater, 2003). Consequently, Hurford (2011) proposes adding a numerical *counter* to finite-state transition networks to accommodate such patterns, suggesting that this amounts to a “significant increase in the power of the processing mechanism” (p. 54).

However, “counting” up to a fixed bound or counting within finite interval is well within the descriptive power of ordinary finite-state transition networks. One simply grafts on a sequence of states that spells out the possible integers from 4 to 11. **Figure 3E** displays a simple illustrative example that captures the “4–11” chaffinch syllable patterns, though it saves space by only displaying a network that counts out four through seven *warble* syllables. The network uses transition arcs labeled with *warbles*, as well as a second kind of transition, labeled with an epsilon, which means that one can move between the indicated states without a corresponding output syllable. In this way, the network can count out four *warbles* and then move to its final state; or five *warbles* and move to the final state, and so forth). This is not the only way to implement finite “counting” bounds of this sort, while remaining within a finite-transition network framework. As is familiar from the literature on finite-state machines, bounded arithmetic operations are straightforward to implement in finite-state devices. Minsky (1967) has many examples illustrating how finite-state adders and counters of this sort may be implemented. In any case, as we describe in more detail just below, such patterns, even of this iterative sort, still form a highly restricted subset of the entire set of patterns that the finite-state transition networks can describe, crucially, one that is easily learned from positive exemplars of adult tutors’ songs to juveniles.

What sorts of constraints *cannot* be described by finite-state transition networks? Roughly, such systems cannot describe containment constraints that can be arbitrarily nested, in the sense that the state transitions generate syllable sequences in form, (*warble*$_1$ (*warble*$_2$ (*warble*$_3$ … *rattle*$_3$) *rattle*$_2$) *rattle*$_1$). Here we have indicated that *particular warbles* and *rattles* must be paired with each other by the use of subscripts, matching from the inside-out, so that the innermost *warble* must be associated with the innermost *rattle*, the next innermost *warble* with the next innermost *rattle*, and so forth. The ellipses indicate that a song might have, at least in principle, an indefinite number of such nestings, to any depth. We have artificially introduced parentheses to more clearly indicate the grouping structure, which is not actually part of the string sequence. Long-standing results (Chomsky, 1956; Rabin and Scott, 1959) demonstrate that such patterns cannot be generated by any finite-state transition network, because, for example, in order ensure that each *warble*$_i$ on the left is matched with its corresponding *rattle*$_i$ on the right one must in effect be able to match up *warbles* and *rattles*, working from the innermost *warble*$_i$ *rattle*$_i$ pair outward. To do this matching requires the machine to use one state to “remember” that an *warble*$_i$ has been seen, until the corresponding *rattle*$_i$ has been seen, one state for each possible *warble*$_i$. But this means that to check a candidate string *warble*$_1$ *warble*$_2$ *warble*$_3$…*warble*$_n$ *rattle*$_n$ *rattle*$_{n-1}$…*rattle*$_2$ *rattle*$_1$ for validity, one must have at least *n* states in the corresponding transition network. If *n* can be arbitrarily large, no machine with a finite number of states will be able to do the job correctly; an indefinitely large memory is required. At a minimum, one must augment a finite-state network with a single counter that is increased by 1 each time a *warble* is seen, and decremented by 1 each time a *rattle* is seen, and the counter must be able to “count” arbitrarily high.

To handle such examples, one must move to a more powerful computational device, such as recursive transition networks (Woods, 1970); equivalently, context-free grammars. For networks, the augmentation involves some means of invoking subportions as though they were subroutines in a computer program. This can be done by expanding the domain of labels on transition arcs to include the names of whole networks, rather than just output symbols such as *warble* or *rattle*. **Figure 3F** illustrates one way to build such a network, where we have numbered the states for convenience. Referring to this figure, we name this entire three-state network with the label *S* and then add a transition from the second state of that network back to that same second state via a transition labeled *S* (the name of the entire network itself). Such a network machine can be organized to use itself as a subroutine, to spell-out all and only the legitimately paired *warble-rattle* sequences.

To see how such an augmented network can generate the syntactically valid string *warble-warble-rattle-rattle* we can again trace through an imagined traversal from the *Start* state to the *Final* state of the network. Again referring to **Figure 3F**, the machine begins in the *Start* state 1, and then travels to state 2, corresponding to *warble*. It can now traverse the network *S* again, by following the loop labeled *S* that goes from state 2 back to state 2, rather than making a transition to the *Final* state (and outputting a *rattle*). This means moving to state 1 again, with the proviso that the network implementation must “remember” that it must return to state 2 when it has traversed the *S* network successfully, by arriving at the final state. We now suppose that during this second passage through the *S* network the machine moves from state 1 to state 2, and outputs another *warble* as before, so that so far the sequence generated is *warble-warble*. If we now have the machine make a transition to state 3, the final state of the network, it adds a *rattle*, which in this case is paired up with the immediately preceding *warble*, as required. However, instead of simply ending its computation at this point, the network has only completed its second traversal of the entire *S* network. It thus must remember that it is required to return to the state where the *S* network was invoked for the second time, namely state 2, and can finish by making a transition from state 2 to state 3, outputting a second *rattle*. In this way the network generates (alternatively, verifies) the desired, legal syllable sequence *warble-warble-rattle-rattle*.

To organize a transition network this way so as to be able to use its parts as if they were subroutines is typically implemented by means of an additional, special memory structure, what is called a *pushdown stack*. As is familiar, a pushdown stack stores information in a first-in, last-out order, like a stack of dinner plates: if items *x*, *y*, and finally *z* are placed on the stack in that order, then the order in which they are removed must be the reverse of this, namely, *z*, *y*, *x*, in this way obeying the characteristic "nested" structure in our example. So for example, traversing the network *S* for the first time, the number of the state to return to, say, 2, would be placed on the pushdown stack. After traversing the *S* network the second time and moving to the final state, the machine would examine the top symbol on its stack, remove it, and returning to the state indicated, in this case state 2, and continue. In this way, a sequence of $n-1$ *warbles* would result in a corresponding sequence of $n-1$ invocations of the network *S* and $n-1$ instances of state symbol 2 being placed on the pushdown stack. Returning from this sequence of invocations in turn and traversing from state 2 to 3 each time will output $n-1$ *rattles*, leaving the machine in state 2 with a single final *rattle* transition to make to reach the end of its very first full traversal through the *S* network, generating the proper sequence of *n warbles* followed by *n rattles*. (As suggested above, since one need only put a single fixed state symbol 2 on the pushdown stack, one could also implement this particular network with a single *counter* that simply indicates the number of 2's that have been placed on the stack, decrementing this counter as each transition to the final state is made.)

Adapting this approach to human language requires more. If we have at least two networks with different labels, say S (corresponding to a Sentence), and NP (corresponding to a Noun Phrase), then the resulting system can be set up to generate Noun Phrases properly containing Sentences, and vice-versa, in the manner suggested by our *the rat chased the birds...* example cited in the Introduction. Such a system would place at least two distinct symbols on its stack, corresponding to the two different types of phrases. This seems to be the minimum augmentation required to describe human language syntax, and goes beyond augmentation of a finite-state transition network with a single counter. One can choose to augment a finite-state device with two counters, but this makes such a machine as powerful as any general-purpose computer (Hopcroft and Ullman, 1979), which would seem to be quite powerful indeed. Below we suggest that human language may be more restricted than this.

It has also sometimes been suggested (see, e.g., Hurford, 2011; Scharff and Petri, 2011) that the shallow hierarchical structure of birdsong, with syllables organized into motifs, and then into some linear motif sequence, could be interpreted as representative of a *general* hierarchical structure-building competence in birds. This conclusion seems too strong. Note that the hierarchical structure here is quite limited. It is comparable to how linguists have described the sound structure of human words in terms of linear syllable sequence "chunks." For example, the word *starling* can be broken down into two consonant (C) vowel (V) combinations, with the first consisting of two consonants, *st-ar* and *l-ing*, that is, CCV–CV. Here the same CCV combination shows up in other words, such as *startle*, so it is similar to a birdsong chunk or motif. We may call this kind of re-use of a linear sequence *linear grouping*. In any particular language such as English, only certain linear groupings are possible. For example, the sequence, *st-ar* is possible in English, while *st-xa* is not. In this way, legitimate CV sequences can be spelled out as allowed linear grouping sequences. This also appears to be true of birdsong.

In both birdsong and human language, this kind of linear grouping has also been shown to have psychologically verifiable correlates. For example, Suge and Okanoya (2010) demonstrated that Bengalese finches perceive songs in terms of syllable "chunks" that can be detected by placing a brief noise either at the boundary of chunks or in the middle of chunks, while training birds under operant conditions to react to the noise as quickly as possible. The birds' reaction time differed in these two conditions, with a longer reaction time for noise introduced into the middle of a chunk, indicating that birds reacted to "chunks" as categorical units for production. In humans, syllable chunks have been found to be an integral part of perception, even in very young infants as early as 4 days old (Bijeljac-Babic et al., 1993).

Some researchers have suggested that linear grouping implies that the underlying birdsong *must* be modeled by a recursive transition network system, but this conclusion too seems unwarranted. For example, Hurford (2011) posits that nightingale song necessitates description in terms of context-free rules (what Hurford calls, "phrase structure rules," equivalent to what augmented transition networks can describe). Hurford further grounds his claim on certain neurophysiological evidence from Fee et al. (2004) regarding the interaction between HVC–RA nuclei in zebra finches' brains during song production. Hurford advances the hypothesis that there is a putative rule expanding a finch birdsong motif as a particular set of seven syllables, *a* through *g*, that is literally represented in a finch's brain by means of HVC–RA interaction, where this rule may be invoked any number of times:

(2) $Motif1 \rightarrow a\ b\ c\ d\ e\ f\ g$

However, it is extremely difficult, if not impossible, to distinguish this possibility from one that simply encodes this sequence as a small finite-state transition network, as displayed in **Figure 3G**. Note that the finite-state transition network, as usual, uses only a small finite amount of memory; it seems entirely possible that a bird could store dozens of such network snippets. No stack-like augmentation is necessary, since, as Hurford himself notes, the entire system in such cases remains a first-order Markov network. By the definition of a first-order Markov system, a finch does *not* have to "remember" whether a motif of one type is "embedded" within another of the same type; it simply has to branch to the part of the network shown in **Figure 3G** at any one of a number of distinct points within a larger, overall song sequence. The sequence would remain entirely linear. Counterfactually, if it were the case that finch song incorporated *nested* dependencies of the "warble-rattle" sort that we described above, then one would be forced to use a more powerful network. But as Hurford himself states, this does not appear to be true of the finch's song. Further, in a recent study, Katahira et al. (2011) demonstrate that very simple first-order Markov processes, even simpler than the "long-distance" finite-state transition networks described above, along with interaction between the HVC–RA bird brain nuclei, can yield apparently "higher order" syllable constraints of precisely sort that Hurford describes.

Currently, then, there is no compelling evidence that recursive transitions networks *must* be literally encoded in finch's brains. To distinguish between the finite-state and non-finite-state possibilities demands artificial language learning experiments that are carefully crafted to distinguish between these two possibilities, along the lines of the experiments carried out with human subjects by Uddén et al. (2011). There is one recent, controversial artificial language learning experiment in Bengalese finches (Abe and Watanabe, 2011) that superficially appears to run counter to this conclusion. However, as demonstrated by Beckers et al. (2012), and as we touch on briefly below, the experimental design here seems to be flawed because the training and testing materials confound acoustic familiarity with syntactic well-formedness. In fact, Uddén and colleagues show that even in the human case, it can be extremely difficult to distinguish experimentally between the use of adjacent dependencies, requiring only a first-order Markov description, and non-adjacent dependencies that might tap the power of a pushdown stack. Absent such careful experimentation, which has to date not been carried out in birds, all current evidence suggests that only finite-state transition networks are required to describe a bird's "knowledge of birdsong." Indeed, it would be surprising if this were not true, since this is in line with what is also known about the acquisition and use of human sound systems as well (Heinz and Idsardi, 2011).

As mentioned earlier, birdsong appears to be much more constrained than this, however. It appears to be describable by a narrowly constrained subset of the regular languages (Berwick et al., 2011a), namely, those that are learnable in a computationally tractable way from examples sung to juvenile males by adult tutors. Here "computationally tractable" adopts its usual meaning in computer science, namely, computable in a length of time proportional to kn, where n is number of states in the to-be-acquired network and k is a small "window size" of one to three syllables. This is an extremely favorable result from the standpoint of both perceptual processing and learning, since in general, learning finite-state transition networks is not possible even given a large number of positive examples, possibly exponential with respect to the number of states in the final, to-be-acquired network (Gold, 1978). Intuitively, this is true of general finite-state transition networks because if all we know is that a target automaton is a finite-state automaton with n states, then it could take a very long string to distinguish that automaton from all other possible n-state machines. More precisely, it appears that one can characterize the formal complexity of birdsong sound systems as a so-called k-reversible finite-state transition network (Angluin, 1982; Berwick and Pilato, 1987; Berwick et al., 2011a). Sasahara et al. (2006) have shown that one can in fact apply the same computer algorithms described by Berwick and Pilato to the problem of automatically inducing k-reversible transition networks from birdsongs. For instance, the finite-state transition network described in **Figure 3B** is k-reversible.

There is no comparable learnability result for human language sentence syntax. However, if one restricts one's domain to human language sound systems, as Heinz (2010) among others have shown, one can obtain a comparable positive learnability result. In this respect then, birdsong and human sound systems again seem alike in terms of ease of learnability (Heinz, 2010). In this context, it should be noted that it is sometimes suggested that the difficulty of learning human syntax as established by Gold (1978) and others can be overcome by adopting another learnability framework. For example, one might adopt a statistical approach, such as rules that apply probabilistically; or a learning system that selects rule systems according to a size criterion (where a smaller rule system is better; equivalently, a Bayesian formulation); While a detailed analysis of such proposals like these lies outside the scope of this paper, in fact while these methods might eventually turn out to be successful, none of them solve the problem of human language acquisition. Such alternatives were originally advanced by Solomonoff (1964), Horning (1969), and later pursued by Berwick (1982, 1985), Stolcke (1994), De Marcken (1995, 1996), and, more recently, Chater and Christiansen (2010) and Hsu et al. (2011), among several others). However, these results have yet led to provably efficient algorithms that cover substantial linguistic knowledge beyond sound systems. Simply making rules probabilistic actually does not work, particularly for sentence syntax that is not describable by means of a finite-transition network. One this point see, e.g., Stolcke (1994), De Marcken (1995), and Niyogi (2006) for further discussion as to why this is so. Intuitively, it is actually more difficult to estimate probability *distributions* over some function that learns a rule system than simply learning the learnability function itself. In particular, current alternative approaches either advance a method that has *no* corresponding constructive algorithm, let alone an efficient one (Solomonoff, 1964; Hsu et al., 2011); or rely on the hand-construction of an initial grammar that is in any case covers but a small fraction of the human language system (Perfors et al., 2010). (See De Marcken, 1996; Niyogi, 2006, for further discussion of why moving to a probabilistic setting does not solve the difficult question of language learnability; and Berwick et al., 2011b for a detailed analysis of recent approaches.)

Formally, a finite-state transition network is k-reversible if, when we exchange the *Start* and final states, and then reverse all the directed transitions from one state to the next, then the resulting new network can be traversed deterministically, that is, without choice points. More intuitively, what this means whenever two prefixes of a song whose last k words match have an end-sequence in common, then they have *all* end-sequences in common. A juvenile learner can acquire such a language by considering examples of an adult male's song, incrementally. For example, if it is the case that a valid song consists of sequences such as *warble-rattle*; *warble-rattle*; *warble-rattle-rattle*; *twitter-rattle*; and *twitter-rattle-rattle*, then all the sequences following *warble* or *twitter* are shared, and the language is 0-reversible. If this hypothetical birdsong language contained in addition the sequence *warble-rattle-rattle-rattle*, since the end-sequence *rattle-rattle-rattle* does not follow *twitter*, then the language is not 0-reversible, unless the bird "generalized" its language to include *twitter-rattle-rattle-rattle*, thereby maintaining 0 reversibility. The 1-reversibility constraint is similar, buts adds an additional syllable of "lookahead," a predictive window 1-syllable long: it asserts that if some syllable *plus* 1 additional syllable – so a chunk of two syllables in all – has *one* suffix in common with another two syllable chunk with the same second syllable, then such a pair of two syllable chunks must have *all* suffixes in common. We can illustrate the difference between 0 and 1-learnability with another caricatured birdsong example. Suppose the set of possible songs consisted of the following five syllable sequences: (1) *warble-rattle-twitter*; (2) *twitter-warble-twitter*; (3) *warble-rattle-tweet*; (4) *warble-trill-tweet*; and (5) *twitter-rattle-tweet*. First, note that *warble* and *twitter* do *not* share all suffixes in common, since in sequence (4) *warble* can be followed by *trill tweet*, but there is no similar suffix for *twitter* – the sequence *twitter-trill-tweet* is *not* part of the song repertoire. Thus, the song language is *not* 0-reversible. However, the language *is* 1-reversible. To test this, we observe which single syllables are held in common between *warble* and *twitter*. There is one such case, for the syllable *rattle*, in sequences (3) and (5), where we have *warble-rattle-tweet* and *twitter-rattle-tweet*. Since in both such sequences (3) and (5) share all suffixes past *rattle* in common, namely, *tweet*, the 1-syllable "window" test is met, and the language is 1-reversible. The extra syllable *warble* makes all the difference. From a learnability stand point, if such a constraint holds for some relatively small value of *k*, then the resulting song is easy to learn just by listening to song examples, as Sasahara et al. (2006) have shown by a direct computer implementation, with *k* at most three.

Taken together then, all these results so far point a single conclusion: birdsong is more closely analogous to human speech than human language syntax. Even so, one must be cautious here as well, because even human speech and birdsong are different from one another in certain respects – unsurprisingly, birdsong is *song*, and human speech does not have all the aspects of song; Fitch (2006) has a thorough review of this comparison. In particular, both birdsong and human songs include as essential aspects both explicit *pauses* and *repetition* – as was noted in the discussion of chaffinch song. One need only bring to mind any Mozart aria to recognize that in human song, even when accompanied by words, pauses, and repetition play a key role in the music itself. This is not typically the case in human speech or language. In language, pauses can indeed be found as part of the descriptive prosodics of an utterance, as in the brief pause after a stressed focal item, as indicated by commas. But pauses are not integrated into the acoustic speech stream in the same essential way as in music, where specific numbers of pauses and pauses of particular lengths *must* occur in certain places, as is clear from musical notation. Repetition is also found in human language, but also strictly delimited, for example, the type that linguists call "reduplication," the repeated occurrence of particular words or morphemes, often indicating some manner of intensification, as in, *very*, *very*, *cold* (see, e.g., Marantz, 1982). Like pauses, the role of particular repetitions in human language is much more limited than in song, where entire phrasal units are deliberately repeated. Putting aside the lack of words, the analogy between human song and birdsong seems in fact extremely close. All things considered, birdsong might serve best as a comparative model for human song, and secondarily for human speech, encompassing vocal learning and vocal production.

HUMAN LANGUAGE AND BIRDSONG: THE KEY DIFFERENCES

As outlined in the Introduction, in human language, hierarchical grouping is also accompanied by additional properties not found in birdsong or human sound systems. Let us revisit these, and then see in Section "A Model for Human Language Syntax" how they might be modeled by a single, very simple combinatorial operation. For reference, **Table 1** brings together in one place the birdsong–human language comparative results described in this section and the article as a whole.

First, human language admits indefinitely extendible, asymmetric containment relationships with at least *two* (and generally more) distinct types of labels. A sentence-like *John knows the starlings* can contain another sentence, as in *John knows the starlings will eat the apples*. Even this possibility does not seem to arise in sound systems, where legal motifs (in the case of birdsong) or legal consonant–vowel possibilities (in the case of language) do not form whole units that are in turn further contained within one other, e.g., we do not find human consonant–vowel structures in the form, (CV(CV(CV))), with the parentheses demarcating the containment of the leftmost consonant–vowel component inside two others.

The multiple types of phrases derive from a second property of natural language structure not found in birdsong, and that is *labeling* dependent on *word features*. The phrase *ate the apples* has the properties of a particular component based on the features a just one lexical item, the verb *eat* (surfacing as *ate* in its past tense form). Note that while it is logically possible to fix the properties of a Verb Phrase in some other way – say, by using the properties of the Noun Phrase *the apples*, or by somehow combining the properties of *ate* and *the apples*, that is not the way human syntactic machinery seems to operate. For this reason, the phrase *ate the apples* is conventionally called a Verb Phrase (VP; rather than a noun-like phrase or something in between). We can say informally that the phrase is *labeled* by *selecting* the verb and certain of the verb's features, and this how the phrase inherits "verb-like" properties. Similarly, a phrase like *the apples* is conventionally called a Noun Phrase (NP). Here we will simply assume informally that the features for the label of this kind of phrase are drawn from some properties of the noun *apples*, namely, that is a noun.

Using the conventional notational system devised by linguists, we can write out the hierarchical structural description for *eat the apples* in a bracketed notation in lieu of the graphical description of **Figure 1A**, where the opening and closing square brackets with labels indicate the extent of a phrase:

(3) $[_{\mathrm{VP}}$ *eat* $[_{\mathrm{NP}}$ *the apples*$]_{\mathrm{NP}}]_{\mathrm{VP}}$

Since the VP label is a simply an arbitrary gloss for particular properties of *eat*, we may replace it with the label *eat** where *eat** denotes these verbal features, whatever they might be. We can do the same for the Noun Phrase, or NP. We further suppress the label on the closing right brackets for readability, arriving at this representation for the syntactic structure corresponding to the sentence:

(4) [$_{eat}$* *eat* [$_{apples}$* *the apples*]]

We now recall that this bracketing structure is different from a linear word or sequence pattern, as in a consonant–vowel combination or a birdsong motif. The key difference is the use of a verb or noun's features to label an *entire* word sequence with a single label, in our gloss, *eat**, or *apples**. As we described in the Introduction, the selection of a privileged element in this way renders the underlying structure fundamentally asymmetric. Note that there is no analog to this in birdsong, a second key difference with human language. Consider as an example the birdsong motif described earlier, consisting of seven particular syllables. This motif is not "labeled" by selecting just one of these syllables and its properties to name the entire motif; none of the syllables takes priority in the same way that *eat* does in the human language example. Neither is the resulting structure asymmetric as it is in human language. This is true precisely because birds apparently do not have words or manipulate word features at all. This is one difference between the human language syntactic system and birdsong. We noted earlier that this does not in principle bar the possibility of birdsong making use of features of song elements, for example, syllables and their acoustic features, and assembling them in a similar hierarchical fashion. However, current evidence suggests that this does not occur in birdsong. Rather, the combinatorial operator itself is absent.

A third difference between human language and birdsong also follows. Once a labeled phrase is available to the human language syntactic engine, it can enter into additional syntactic manipulations as a new, single unit, as if it were a single word. So for example, once having established *eat the apples* as "chunk" *eat**, the human language system uses *eat** as a single verb-like object to build forms such as, *the starlings will eat**, i.e., *the starlings will eat the apples*. More interestingly, even more complex examples with *eat** can be constructed, such as, *the starlings will eat the apples and eat the apples the starlings did*, where *eat the apples* is understood as occurring in at least three different places: (1) at the start of the sentence; (2) after *and*; and, more interestingly, (3) in an unpronounced (phonologically null) "understood" form after *did* that is interpreted in exactly the same way as if *eat the apples* was actually present after *did*. More precisely, one can say that *eat the apples* is in fact present in the *syntactic structure* following *did*, does not surface phonologically – that is, it is not spoken or signed. This happens when the internalized syntactic form must be externalized; the third occurrence of *eat the apples* is suppressed and remains unrealized as part of the sound/manually signed stream.

This last example is quite characteristic of human language as we shall see with additional examples. However, it is absent from birdsong, where there are no "unpronounced" song components, virtually by definition. If we let *eat** denote the label of the entire phrase, *eat the apples*, then we can write out the syntactic structure of this last example as follows, where *S* denotes a sentence, and we have suppressed irrelevant details, like the analysis of *will* and *did*, that carry tense:

(5) [$_{S}$ [$_{starlings}$* *the starlings*] [*will eat**] *and* [$_{S}$ *eat** [$_{starlings}$* *the starlings*] [*did eat**]]]

We can see in this example that the syntactic structure has encoded a dependency between these three occurrences of *eat**: they are in effect linked copies, in the sense that they refer to the same syntactic object, *eat the apples*, but the copies appear in several different positions in the sentence. In the same way, given the sentence, *the starlings ate the apples*, if we label the phrase *the apples* as, *apples**, then one can form a sentences such as, *the apples the starlings ate*, which is interpreted as, *apples** *the starlings ate apples**. In this case, *the* apples is interpreted in two positions. The first position is at the front of the sentence, corresponding to its role as the so-called "topic" or "focus" of the sentence (which carries a special intonation peak, as the comma indicates). The second position is as the Noun Phrase adjacent to the verb *ate*, corresponding to the status of *the apples* as the Object of the verb, just as in the simple declarative sentence, *the starlings ate the apples.* The fact that one and the same phrase can be, indeed must be, interpreted in two distinct places in a sentence, one associated with discourse factors, and the second, with semantic interpretation as the argument of a predicate, is yet another wide-spread phenomenon in human language, absent in birdsong. This kind of "displacement" of phrases, no matter how it is described, seems nearly ubiquitous in human language, as most standard linguistic accounts note (see, e.g., Radford, 1997).

Birdsong, even when described via "chunks" that might correspond to phrases, does not seem to have *any* of these additional distinctive properties of human language. Let us see in detail why not. If a birdsong motif is made up of, say, two syllable sounds, we do not find that the features of one of these syllables is differentially selected to characterize the motif as whole. This would amount to a representation something like the following, where "warble" and "rattle" are presume to be two distinct birdsong motifs, along the lines of *eat the apples*:

(6) [$_{warble}$* *warble-rattle*]

However, nothing like this seems to be found in birdsong. Nor do we find the embedding of one motif inside another, or the embedding of two *different* kinds of phrases within one another, like a Sentence within a Noun Phrase. Finally, we do not find examples like *eat the starlings did eat* [*the apples*], with unpronounced, but elliptically understood syllabic chunks. In short, *none* of these distinctive properties of human language that move it beyond the domain of simple linear sequencing seem to be found in birdsong.

To indicate how crucial and wide-spread this difference is, we will describe several more such examples of what we might call "copy dependencies," all absent in birdsong, but present in human language syntax. First consider the example below, where there are *two* positions that seem to contain unpronounced "copies" of a Noun Phrase:

(7) *this is the bird that the starlings saw without leaving*

This is a rather more complex sentence. In this example, *the bird* serves as a phonologically suppressed copy in two places: it is the Object of *saw* and it is the Object of *leave.* We can "reconstruct" the unpronounced form *the bird* in these two positions to recover the required structure for proper semantic interpretation (though the sentence sounds more unnatural after this reconstruction):

(8) *this is the bird that the starlings saw the bird without leaving the bird*

Second, one can observe that there are examples where multiple dependencies can be "nested," precisely as in our example of **Figure 1E**, corresponding to the sentence, *the rat chased the birds that saw the cat that ate the starling.* Referring back to that figure, note that we had left empty parentheses for the Noun Phrases that serve as the Subject of both *saw* and *ate.* It is clearly the case that *the birds* is the Subject of *saw*, (e.g., it is the birds that saw the cat) and *the cat* is the Subject of *ate.* We can now clarify that these positions actually represent the same sort of unpronounced, phonologically null instances as in our earlier examples, so that the corresponding Noun Phrase in each case may be replaced in the now reconstructed syntactic form, again so that proper semantic interpretation can proceed. That is, the reconstructed sentence is something like the following, using now the labeled bracket notation:

(9) [$_{\text{S}}$ [$_{\text{NP}}$ *the rat*] [$_{\text{VP}}$ [$_{\text{V}}$ *chased*] [$_{\text{NP}}$ [$_{\text{NP}}$ *the birds*] [$_{\text{S}}$ *that* [$_{\text{S}}$ [$_{\text{NP}}$ *the birds*] [$_{\text{VP}}$ [$_{\text{V}}$ *saw*][$_{\text{NP}}$ [$_{\text{NP}}$ *the cat*] [$_{\text{S}}$ *that* [$_{\text{S}}$ [$_{\text{NP}}$ *the cat*] [$_{\text{VP}}$ [$_{\text{V}}$ *ate*] [$_{\text{NP}}$ *the starlings*]]]]]]]]]]]

Again referring to **Figure 1E**, it is evident that we now have at least *two* sets of dependencies: between *the cat* and its unpronounced position; and between *the rat* and its unpronounced position. Furthermore, crucially these dependencies are "nested" and could be arbitrarily extended in the manner discussed in Section "Birdsong Seems Analogous to Speech, Not Syntax." As we have seen, these kinds of patterns cannot be captured by any finite-state transition network.

As a final example of a more complex dependency found in human language syntax, there are examples that involve what are called *crossed-serial dependencies* as opposed to *nested* dependencies. Such dependencies are called *crossed* because the relationship between the elements overlap rather than nest (see **Figure 4** for an example). These are evidently less common in human language sentence syntax. Among the first examples were described by Huybregts (1984) in certain Dutch and Germanic dialects. But even in English, examples of such dependencies can be found in circumscribed contexts. The classic example was provided by Chomsky (1957), to account for the sequence of English auxiliary verbs and their morphological endings indicating aspects of tense, such as the passive or perfective endings of a verb. Chomsky noted that the apparent underlying syntactic form of the auxiliary verb sequence that is pronounced as, e.g., *will have been being eaten* is best described by the following context-free rule. (See Lasnik, 2000 for additional discussion of this point; we roughly follow Lasnik's discussion and notation below).

(10) Verbal "motif" → Tense-element Modal-verb-∅ *have en be ing be en eat*

That is, to generate English verb auxiliary sequences, the *en* suffix that follows *eat*, to form *eaten*, indicating a passive inflection, is actually attached to the preceding auxiliary verb *be*. Similarly, the suffix *ing* following the last *be*, to form *being*, indicating progressive tense, is actually attached to the *be* that precedes it. The pattern continues all the way through: the suffix *en* that follows the first occurrence of *be*, forming *being*, is actually attached to *have*; and, finally, the "zero" suffix after *have* is actually attached to the modal-verb *will.* If we then draw out these relationships, as shown in **Figure 4**, it is clear that the dependencies between the elements follow a crossed-serial pattern.

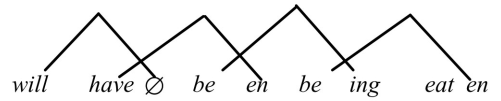


FIGURE 4 | The "crossed-serial" dependency structure of the English auxiliary verb system. Each auxiliary verb, *will, have, be*, and *be*, is linked to the suffix of the verb that follows. The zero element associated with *will* is a so-called "zero morpheme" – the suffix is present, but not pronounced.

Such examples have important computational implications: they require even more sophisticated networks (or grammars) than those we have yet described. Informally, this additional power amounts to having the individual memory locations in a pushdown stack themselves act like separate stacks, alternatively, to have a second, additional pushdown stack. A variety of related formalisms for processing such patterns have been proposed (see, e.g., Weir, 1988), under the name of *mildly context sensitive languages* (and their corresponding grammars). The second stack-like memory behavior is required because in the case of overlapping or crossing dependencies, one must be able to retrieve and insert whole phrases at positions other than those that occur in a last-in, first-out order.

Labeling-plus-grouping also imbues human language syntax with two final characteristic properties that do not appear to be found in birdsong. The infiltration of word features into language's syntactic system serves as a key "hook" between the conceptual atoms underpinning individual words and indefinitely large sentence structures, yielding the open-ended conceptual character of human language generally. This follows from the principle of compositionality grounded on syntactic structure, originally formulated by Frege, as noted by Fodor (1996): if one has separately acquired the words associated with *apples, bananas*, etc., along with the verbs *eat* and *want*, then the algebraic closure of the grouping and labeling operation implicit in forming *eat the apples* applied to this miniature lexicon yields the cross-product of the two possibilities, *eat apples*, . . ., *eat bananas, want apples*, . . .,*want bananas.* In short, we immediately obtain an account of the open-ended productivity of human language as a side-effect of syntactic combination, along with a link to conceptual productivity.

Summarizing, **Table 1** lays out all the comparisons between birdsong and human language that we have surveyed in this article. There are just two areas where birdsong and human language align; this is between birdsong and human language sound systems. All other areas differ. Considering these differences, along with the current evidence from analogous brain regions, to genomics, to vocal learning and production, to the extent that birdsong and, human sound systems are comparable, they align at one particular formal level, that of "input–output" externalization systems, including that of sensory–motor-driven vocal learning. While this is by no means an insignificant part of the entire language system, it is the part devoted only to externalization, and even then this does not address the modality-independence of human language, which can be non-acoustic in the case of manual signing. In this sense, a large gap remains between human language syntax proper and birdsong.

In the next section we outline how a human-capable syntactic system, evidently then quite different from that of birdsong, can be modeled.

A MODEL FOR HUMAN LANGUAGE SYNTAX

All the distinguishing properties of human language listed in **Table 1** can be modeled within a simple computational system, first developed in Chomsky (1995) and in many subsequent places by others. The model described below is not intended to cover all aspects of human language syntax, but rather delimit a minimum set of assumptions, shared by many different linguistic theories, that account for the birdsong–human language syntax distinctions of **Table 1**.

First, whatever other characteristics such an algorithm must have, above all it must be able to associate unboundedly many strings of words with structured expressions, of the sort noted in **Table 1**. Familiar results from the study of computational systems since the latter part of the twentieth century have shown that any such system requires some kind of primitive combinatory operation that can construct larger objects from smaller ones, where the smaller objects may themselves be complex, but ultimately reduce to some set of atomic items, in our case, essentially words (Kleene, 1953). The required combinatory operation has been cast in many forms both in abstract computational systems and in specific generative proposals for describing human language syntax. For example, one such system is the Lambek (1958) calculus, in which individual words have properties corresponding to the "valences" of chemical theory, dictating how they allowably combine with other words or structures. For example, in this system, *ate* has the property (NP§)/NP, meaning that it requires an NP Subject to its left, and an NP Object to its right. Further, there is a single rule of combination that "glues together" two words or previously assembled structures into larger wholes, ultimately an entire sentence. For instance, given *ate* with its properties as above, and a corresponding NP to its right, associated with, say, *the starlings*, the Lambek combinatory operator takes as input these two items, (NP§)/NP and NP, and output a new structure, NP§, corresponding to a traditional Verb Phrase, with the "NP§" notation indicating that a Subject NP is still required to the left. See, e.g., Steedman (2000) for a broader and more recent treatment of human language syntax within an extended version of this framework.

In a similar spirit, here we will also assume a combinatorial operator that associates strings of words with structures, along with a rule of combination, though of a different sort, along the lines described by Chomsky (1995), where this combinatorial operator is called "merge." We follow Chomsky's presentation closely in what follows below, because along many dimensions it is relatively theory neutral, in the sense that it makes the fewest possible assumptions about the syntactic machinery needed to generate possible sentence structures – many current linguistic theories contain at their heart some sort of combinatorial operation similar to the one described here. Also like the Lambek calculus, the inputs X and Y are either individual lexical items, what we will also call *atomic* units, or else more complex syntactic objects previously constructed by application of the operator from such atomic units. Finally, also in accord with the Lambek system, we assume that the combinatorial operator can apply to its own output, that is, a previous application of the combinatory operation. Following the Church-Turing thesis, this is a requirement for yielding a system that can associate indefinitely many words with structured syntactic objects.

At this point, our model diverges from the Lambek system. Unlike the Lambek calculus, for computational simplicity, following Chomsky (1995), we will assume that X and Y are unchanged by the combinatorial operation, so we can represent the output simply as the *set* $\{X, Y\}$. Assuming otherwise would take additional computational effort. For example, if $X =$ the single lexical item *ate*, and $Y =$ the more complex syntactic object corresponding to the phrase *the starlings*, then the output from the operator given this X, Y input would simply be the set, $\{ate, Y\}$. Referring to more traditional notation, this particular set would correspond to what we earlier called a Verb Phrase, with X equal to the atomic item *ate*, and Y equal to the Noun Phrase Object associated with *the starlings*.

In further contrast with the Lambek system, note that this output set is by definition unordered, in this way reflecting the apparent lack of any syntactic predicates based on linear precedence. Any such order is assumed to be imposed the sound system of the language, which, for example, must determine whether verbs precede or follow Objects. (English chooses to externalize syntactic structure so that Objects follow Verbs, while in German or Japanese the choice might be otherwise.) Crucially, the operator can apply again to its own output, so generating a countable, discrete infinity of possible syntactic structures (Berwick, 2011).

We will introduce one more bit of machinery to describe this system, and that is the notion of *labeling*. Here too, as in the Lambek system, after the combinatorial operation has been applied, the newly created syntactic object has properties that are based on, but not exactly the same as, the properties of the objects out of which it has been constructed. In fact, in the Lambek system, the new composite object is some subset of just one of the features of the two objects that were joined together; in our example above for instance, the structure corresponding to a Verb Phrase, (NP§), obtains its properties from that of the verb, (NP§/NP). We will follow something along these lines, though somewhat distinct. With this notion in hand, now proceed to show how the syntactic structure for an entire sentence, *the birds ate the starlings*, might be generated.

Note that where we have Y identified with an entire phrase, *the starlings*, it must be the case that this syntactic object Y was itself constructed by some previous application of the operator. In particular, we must have applied it to the two lexical items, *the* and *starlings*, so obtaining Y. Recalling our earlier discussion where we described Verb Phrases as inheriting their properties from the properties of verbs, we need in addition a way to identify and label such newly minted sets. To do this we will assume that when applying the operator to two sets X and Y, we must always select the properties of just *one* of these to serve as the *label* for the set combination that results. So for example, when we apply the operator to two sets consisting of simply lexical items, say {*the*} and {*starlings*}, then we select one of these, here *starlings*, and write the actual combination as, {*label*, {*X*, *Y*}}. Following the lead of our example from the previous selection, we gloss these label features as *starlings**. The output result is the more complex syntactic object in (11) below:

(11) *Y* = {*starlings**, {{*the*}, {*starlings*}}}

It is this set that corresponds to the conventional notion of a Noun Phrase, though it is important to recall again the crucial difference that unlike a traditional Noun Phrase, there is no linear order imposed between *the* and *starlings*, which is of no concern to the internal syntactic system. The ordering between these two words is left for the phonological sound system to spell-out, when the phrase is actually pronounced.

With this elaboration in mind, when the operator is applied to inputs *X* = {*ate*}, and *Y* is as defined just above, the syntactic object corresponding to *the starlings*, one must again select a new label for the output of the combinatorial operation. In this case, we assume to be the label drawn from *X*, namely, *ate**. (We leave to one side the question of why it is *X* rather than *Y* that is selected for fixing the label.) Following through with our example then, the operator applies to the two sets *X* and *Y*, yielding the more complex structure in (12):

(12) {*ate**, {{*ate*}, {*starlings**, {{*the*}, {*starlings*}}}}}

This set corresponds to a conventional Verb Phrase, though again without any linear precedence ordering between what would conventionally be called the Verb and the Object Noun Phrase. Finally, by using this set-structure along with the set-structure corresponding to the Subject Noun Phrase, e.g., *the birds*, we may apply the operator once again, outputting a final Sentence structure along these lines:

(13) {*ate**, {*birds**, {{*the*}, {*birds*}}}, {*ate**, {{*ate*}, {*starlings**, {{*the*}, {*starlings*}}}}}}

While this notation appears complex, it in fact contains all the hierarchical information needed to recover the Subject and Object relations, the adjacency of the Object NP with the verb, and in fact any other required syntactic relationships associated with the sentence. Let us see how this works out. Consider the required adjacency relationships. First, the Object must be adjacent to the verb. This is true in our output structure, because in the form:

(14) {{*ate*}, {*starlings**, {{*the*}, {*starlings*}}}}

we can see that {*ate*} and {*starlings**...} correspond to pairs {*X*, *Y*} at the same level in structure (13), and thus meet the correct notion of "adjacent to" required. Note that this property crucially follows because we have (tacitly) assumed that composition always takes two arguments. This is not a necessary property, but one that seems empirically sufficient, as noted in Chomsky (1995). Similarly, the Subject must be adjacent to the syntactic object that denotes the conventional Verb Phrase, and here too we find that the set construction properly describes this relationship:

(15) {*birds**, {{*the*}, {*birds*}}}, {*ate**, ...}

Here, {*birds**, ...} and {*ate**, ...} are pairs *X*, *Y* at the same level, and so adjacent to one another.

Turning to hierarchical relationships, the Verb–Object combination is set off as a phrase distinct from the Subject, in virtue of its *containment* within a subset of its own, apart from the one that contains the set associated with *the birds*:

(16) {*ate**, {{*ate*}, {*starlings**, {{*the*}, {*starlings*}}}}}

Further, the asymmetry of the set-structure is fixed by the very definition of labeling, since only one lexical item participates in determining a label's features.

What is the advantage of this alternative system? Importantly, such a system automatically admits the possibility of examples such as *the birds will eat the starlings and eat the starlings the birds did* [*eat the starlings*], because the combinatorial operator applies to *any* two sets *X*, *Y*, even when *Y* happens to be a subset of *X*. Suppose for instance that we have already constructed a (slightly different) Sentence along the lines of our other example sentence above, corresponding to the syntactic structure for *the birds will eat the starlings*, where we have suppressed certain of the linguistic details for expository purposes:

(18) {*will**, {*birds**, {{*the*}, {*birds*}}}, {*will*, {{*will*}, {*eat**, {{*eat*}, {*starlings**, {{*the*}, {*starlings*}}}}}}}}

Given the general combinatorial operator, one of its choices is to freely select to combine the *entire* set object above as its choice for *X*, along with any proper *subset* of this set as its second choice for *Y*, for example, {*starlings**, {{*the*}, {*starlings*}}} (=11), corresponding to the Noun Phrase *the starlings.* Given this choice for *Y*, the output from the combinatorial operator acting on the pair *X*, *Y*, and selecting the label associated with *X* for the output structure, yields an apparent *copy* of the set representation for *the starlings* as suggested in Section "A Model for Human Language Syntax," shown below in (19). In other words, set-structure for *the starlings* now appears in two places: the first position, that of the "discourse focus"; and the second position as the argument of the verb predicate *eat.* In (19) we have highlighted these two occurrences in bold font.

(19) {*will**, **{{{*starlings**, {{*the*}**, {***starlings***}}}, {*will**, {*birds**, {{*the*}, {*birds*}}}{*will**, {{*will*}, {*eat**, {{*eat*}, {***starlings****, **{{*the*}, {*starlings*}}}**}}}}}}}

When this syntactic structure is sent to the phonological system for output, the second occurrence is suppressed, which we indicate below by striking a line through it:

(20) {*will**, **{{{*starlings**, {{*the*}, {*starlings*}}}**, {*will**, {*birds**, {{*the*}, {*birds*}}}{*will**, {{*will*}, {*eat**, {{*eat*}, ~~{**starlings***, {{**the**}, {**starlings**}}}~~}}}}}}}}

The sound system that "externalizes" this internal syntactic structure will as usual output only the actual words in brackets, not the labels or the rest of the syntactic form, imposing precedence relations, so that the output from the sound system surfaces as (21):

(21) *the starlings the birds will eat*

In this way, a single combinatorial operator, without any additional assumptions, automatically generates the syntactic structures described above in Section "A Model for Human Language Syntax," with copies that are present in at least two places, but that remain unpronounced when the internal syntactic form is mapped into its phonological counterpart and spoken (or manually signed). Furthermore, this way of constructing the form automatically ensures that the discourse prominent copy is hierarchically superior to the copy that serves as the verb's argument, as required. We do not have to specify some new, separate operation apart from the single combinatorial operator in order to generate structures with copies. This is part of the way the operator works with all syntactic objects.

A second major advantage of this system is that it can account for a wide range of syntactic phenomena within both English and across many dozens of other languages where there are apparent "pairings" between forms such as the following:

(22a) The starlings will eat the birds
(22b) Will the starlings eat the birds

Here, as discussed in Berwick and Chomsky (2011), the auxiliary verb *will* in the question form (22b) must be interpreted in the same position as it is the declarative form (22a), in order that the close link between the sense of (22a) and (22b) can be maintained. (The second is the interrogative form of the first.) This property is automatically accommodated under a model where (22b) is formed by the combinatorial operation acting on *will* as a subset of the larger set-structure corresponding to *will eat the birds*. Just as before, an apparent copy of *will* is placed at the end of the sentence, with *will* remaining in its "original" position, where we have inserted brackets to highlight the phrasal and subset-set relationships:

(23) [Will] the starlings [[will] eat the birds]

Once again, when pronounced, the second occurrence of *will* is suppressed, and the sound system outputs the form (22b).

(24a) [Will] the starlings [~~[will]~~ eat the birds]
(24b) Will the starlings [eat the birds]

There is a large range of similar cases that have been investigated by linguists over the past 60 years covering many dozens of languages, all of which can be accounted for by the combinatorial operator posited above. This provides substantial empirical support for the particular assumptions we have made; see, e.g., Radford (1997) among other recent texts for details.

THE EVOLUTIONARY PICTURE

We have framed the "gap" between birdsong and human language in Sections "Human Language and Birdsong: The Key Differences" and "A Model for Human Language Syntax" in way that lends itself to two main evolutionary questions. The first concerns the combinatorial operator itself. Is this computational competence present in other species? If not, how did it arise? Does it have antecedents in terms of older or related competences? Can we break down the operator into smaller components, and use these to envision an evolutionary scenario such that the operator might have been acquired in distinct stages? The second question concerns the stock of atomic elements, the lexical items or words that feed the combinatorial operator. Several possible evolutionary scenarios have been envisioned regarding these puzzles, for the most part difficult to verify, given the absence of the relevant evidence.

We review just one position here: that in fact there is no such gap, and that songbirds (and other non-human species) actually possess the same syntactic combinatorial ability as humans, though lacking lexical items. To determine whether this is so, in recent years researchers have attempted to determine whether songbirds can succeed at artificial language learning tasks. Following the lead of experiments carried out with non-human primates (Fitch and Hauser, 2004), these approaches have most often attempted to probe whether songbirds can learn to discriminate the strings of languages that are not describable by any finite-state transition network. In particular, researchers have focused on artificial languages of the form $a_i f_i$, with any number of matching *a*'s and *f*s. Some experiments have added a distinguished center marker, *c*, yielding languages in the form, $a_i c f_i$. As noted earlier, such a language can only be successfully generated or recognized by a finite-state automaton if it is augmented with a single counter, or, equivalently, adding a pushdown stack with a single symbol. In this sense, it is perhaps the "simplest" example of a language that cannot be recognized by an unadorned finite-state transition network, as noted by Rogers and Pullum (2011). The general experimental methodology is to train subjects on a set of familiarization strings drawn from a language known to be non-regular, and then test the subjects to see if they correctly accept syntactically well-formed examples of the language, and reject syntactically ill-formed examples.

Given success in this task, the implication is that the subjects have acquired and then used a system of rules that go beyond the power of finite-state transition networks. Some researchers have in addition suggested that success in this task implies that the subjects have acquired and then used a particular *kind* of finite-state augmentation, either a rule system equivalent to a fully recursive transition network with a pushdown stack as described above, or, a rule system equivalent to this. It can be difficult to test such details about implementation, even in human subjects, as attested by the recent work by Uddén et al. (2011). Using an artificial language learning paradigm, they found experimental support for pushdown stack storage in human subjects to be lacking. However, they did find experimental evidence that crossed-serial dependencies required additional computational effort, in line with a full two-pushdown stack model mentioned earlier. However, it should be noted that as soon as one imputes a full two-pushdown stack system to a computational device, then this computational machinery is as powerful as any general-purpose computer, i.e., it is as powerful as a Turing machine. It remains unclear how such a device is actually implemented in the brain.

In any case, the experiments attempting to demonstrate that non-human species have a cognitive competence that could be emulated by even a single pushdown stack have so far proved inconclusive. In the first experiment to report an apparent success with this kind of protocol in any bird species, Gentner et al. (2006) used operant conditioning to train and then test starlings on an artificial language defined over an alphabet of acoustically distinct *whistle*, *warble*, *rattle*, and *high-frequency* motifs, drawn from the song of one male starling. Eight distinct *warble* and *rattle* motifs were used to formulate a training language consisting of four-syllable strings in the form, *rattle*$_i$-*rattle*$_j$-*warble*$_l$-*warble*$_j$, with i, j, k, l ranging from 1 to 8, corresponding to a sample from the "correct" target language in the form, $a^i f^i$. This was used for positive-reinforcement operant conditioning. The starlings were also trained to avoid syntactically ill-formed strings of the form, *rattle*$_i$-*warble*$_j$-*rattle*$_l$-*warble*$_j$, corresponding to the language $(af)^i$, a language that can be generated by a finite-state transition network. After many thousands of positive and negative reinforcement trials, the birds were then probed with different novel correct and incorrect sequences, including longer length-6 and length 8-strings, and responded positively to the correct strings while also properly avoiding the incorrect ones. Can one therefore conclude that starlings can acquire and use the rules for hierarchical structures along the lines of human languages?

The answer seems to be no, for at least two reasons. First, the language that was used to exemplify the use of a finite-state transition network with recursive subroutines, alternatively a context-free grammar, was not in fact of the right type to *unambiguously* demonstrate the conclusion that was sought, as noted by Corballis (2007), Friederici and Brauer (2009), and Friederici et al. (2011), among others. Above and in **Figure 3D** we indicated that in such a language the *a*'s and *f*'s must be nested and paired with each other from the inside-out. But this was not true of the artificial language in the Gentner et al. (2006) experiment, where the *warbles* and *rattles* could be of *different* types, not necessarily paired with each other. That is, instead of the language containing strings such as, $a_1 a_2 f_1 f_2$, the starlings were trained on strings including, $a_1 a_2 f_2 f_3$, with the critical nesting property violated. As a result, all that is required for the starlings to succeed on novel, well-formed probe stimuli is that the birds be able to count that the number of *warbles* is followed by the same number of *rattles*. This can be done by a finite-state network with a single, limited counter – that is, all the birds are required to do is to count – subitize – up to this numerical limit, an ability that has already been attested in this species by Dehaene (1997). It is therefore more parsimonious to assume that the birds are simply drawing on abilities that have already been demonstrated, rather than some novel cognitive ability. Second, as suggested by Rogers and Pullum (2011), testing that the starlings reject the "illicit" strings of length six, e.g., *warble-rattle-warble-rattle-warble-rattle* is confounded with the possibility that such strings can also be generated by a non-finite-state transition network, one that tests, in general, whether the length of all the *a*'s is the same as the length of all the *f*s; this is not a language that can be generated by a finite-state transition network.

In part to remedy the first problem, more recently, Abe and Watanabe (2011) carried out an artificial language learning experiment with Bengalese finches, *Lonchura striata* var. *domestica*, concluding that the finches acquired and then used context-free grammar rules for language discrimination. Watanabe and Abe exposed finches to training examples of distinctive song syllables. Birds were exposed to two sets of familiarization strings (denoted FAM), $a_x\ c_z\ f_x$ ("non-embedded strings") and $a_x\ a_y\ c_z\ f_y\ f_x$ ("center-embedded strings," or "CES"), where the letters denote syllables, and matching subscript letters denote matching syllables that always co-occur in a string. These sequences were designed to follow the possible patterns generated by a context-free grammar with the syllables of similar types properly paired. One other difference from the Gentner training language was that in each pattern, the c_k syllable type marked the middle of a legitimate pattern. We can write out an example CES string $a_1\ a_2\ c_3\ f_2\ f_1$ to display the implied containment relationships using bracketing notation as follows, where we have arbitrarily labeled the left-hand square brackets with S_1, S_2, and S_3.

$$[S_1 a_1\ [S_2 a_2\ [S_3 c_1]\ f_2]\ f_1]$$

Watanabe and Abe improved on the Gentner protocol in at least one other respect: no operant conditioning was needed, as the birds' natural calling behavior was used as a response measure.

The finches were then tested to see whether they would reject ill-formed examples such as $a_2 a_1 c_1 f_2 f_1$ (where the order of syllables does not follow the proper nested containment pattern); reject examples like $a_1 f_2 a_2 c_1 f_2 f_1$, where an *f* precedes the *c* marker; and accept as well-formed novel examples such as $a_2 a_1 c_3 f_1 f_2$, where the particular pattern with the center marker c_3 was not part of their training set. The litmus test for recognition (conversely, rejection or non-recognition) was a measurable increase (conversely, a decrease) in calling rate response to the test examples. The finches did vary their calling rates as predicted: calling rates were higher for syntactically correct syllable strings, as opposed to syntactically incorrect syllable strings. At first glance then, this result would seem to confirm that the finches had acquired the syntactic rules for generating nested hierarchical structure, since both the recognition and rejection tasks that would seem to require the grouping of syllables in a nested, hierarchical way.

However, the conclusion that the birds were actually constructing hierarchical representations remains arguable (Beckers et al., 2012). The training and test stimuli were not balanced for acoustic similarity. For example, the correctly center-embedded syllable test strings (COR), e.g., $a_1 a_2 c_1 f_2 f_1$, were largely similar to the familiarization strings, e.g., $a_1 a_2 c_2 f_2 f_1$, $a_1 a_2 c_3 f_2 f_1$, and $a_1 a_2 c_4 f_2 f_1$, both in terms of syllable positions and acoustically, mismatching on just a single syllable, the distinguished center syllable c_i. Thus, even though all COR strings are novel, four out of five syllable positions match in the sense that they contain the same sounds at the same positions. The other group of syntactically correct test strings had a similar problem. This means that the birds could have treated these novel test strings as familiar simply on the basis of their phonetic characteristics alone, without every analyzing their syntactic structure. Since it is already known that Bengalese finches can distinguish a set of familiarization strings as in the Watanabe and Abe experiment as distinct in terms of memorization alone (Ikebuchi and Okanoya, 2000), by this argument we do not need to posit any novel syntactic ability for finches, a more parsimonious explanation for the finches' behavior since it requires fewer assumptions.

Even if similarity matching had been controlled for by using a different experimental design that eliminated the overlap between familiarization and test strings, the sound sequences and presumed underlying structures in this experiment are unnatural in the sense that they are perfectly *symmetric*: there are an equal number of *a*'s, *f*'s, etc. to be matched up on either side of a distinguished center marker. This kind of structure quite unlike the *asymmetric* structures found in natural language, illustrated in **Figures 1A,B**. In fact, even humans have great difficulty mastering artificial languages whose underlying structures are symmetric. At least since the work of Miller and Isard (1964) it has been known that people have great difficulty parsing both naturally occurring self-embedded sentences as well as center-embedded sentences constructed in artificial language learning experiments (Miller and Chomsky, 1963). Confirming this, as briefly described in the introduction it has also long been known that people restructure sentences so as to avoid producing complex center-embedded structures, as well spontaneously using alternative strategies for solving tasks that would otherwise provide evidence for the processing of such structures (Langendoen, 1975, Perruchet and Rey, 2005).

Given the experimental evidence that there is a computational gap in processing ability that reflects a difference between songbirds and humans, then one way to express this distinction is in the way memory is organized. The ability to assemble sequences into groups with distinguishable labels like "VP" (or more carefully, *eat**) and then set them aside for additional later processing suggests the existence of memory locations where these newly assembled units like "VP" might be located and then re-used. At this point, how such memory might be organized remains challenging to discern, given current information about how neuronal structure might "implement" the computational architectures computer scientists are familiar with, as we review briefly below; for a recent discussion and speculations as to several possible "implementations," see Uddén et al. (2011). One problem is that once one moves to a machine with two stacks, one can easily show (Minsky, 1967) that an equivalent computational power can be attained by a machine with just two counters (see also Sipser, 1997 for a good introduction to these issues and automata theory). Such an abstract device could have many possible physical realizations, and at present the empirical evidence under-constrains these.

Pushdown stack storage is sometimes assumed to be implemented by, for example, a set of decaying neural networks, with each offset in a net's decay time corresponding to a different stack location (for more details see, e.g., Pulvermüller, 1993; Pulvermüller and Knoblauch, 2009). Alternatively, Uddén et al. (2011) suggest that arithmetical operations could be used to simulate stack-like operations – one could use a number that grows or shrinks in size, which as they observe might have some more straightforward realization in neural "wetware" than decaying or reverberating circuits. But all such statements should be treated with some caution, because there are many ways of implementing computational devices, particularly if memory access is carried out not according to some symbolic addressing scheme, as in conventional digital computers, but in terms of so-called content-addressable memory. Although algorithms for content-addressable memory are less well-known, even here, hierarchical representations can be readily developed, as described by, e.g., Oldfield et al. (1987). Thus it would be simply incorrect to state by fiat that a content-addressable memory would not be compatible with the efficient storage and manipulation of hierarchical or "tree" structures, even of a fairly complex sort. In any case, from the earliest studies carried by Bever (1970) and Chomsky and Miller (1963), and as further described by Berwick and Weinberg (1985), syntactic structures seem to form locally coherent trees that are then rapidly dispatched for semantic interpretation, so lessening any short-term, local memory load. It remains to explore how a system more faithful to what we know about neuronal structure and operation actually would work to implement this kind of abstract computation.

Earlier we noted that Hurford (2011) suggests that in songbirds, some arrangement of links between the HVC and RA nuclei encode phrase structure rules, but that this analysis is flawed and cannot actually distinguish between finite-state transition network and augmented transition networks. As for non-human antecedents or possible pre-adaptations for pushdown stack storage, it has been proposed that the requirement for this kind of auxiliary storage may have been driven by the requirements for animal navigation such as map following or foraging (Bartlett and Kazakov, 2005; Okanoya, 2007; Shettleworth, 2010). According to these proposals, if animals must remember the particular places where food has been cached, in some particular order, then this might require landmarks to be traversed in the manner of a pushdown stack to retrieve food left at these locations, starting with the most recently visited location first, then the second-to-last visited location, and so forth. As stated, this would amount to a symmetrical visiting pattern, like the embedded strings tested in the Gentner et al. (2006) experiment, in the pattern $a_2 a_1 c_3 f_1 f_2$. While a pushdown stack memory would seem of obvious benefits here, such suggestions have again remained at the level of simulation.

CONCLUSION

What can we conclude about birdsong, human language, and evolution? Birdsong seems most comparable to the sound system of human languages, that is, the externalization of human language proper, encompassing both speech and the manual modality of signed language. This comparison seems to hold both at a formal level of analysis, that is, systems compared as a sequence of acoustic (or gestured) strings, as well as at many distinct levels of neurological analysis, from brain regions down to the genomic level. Given the long divergence time between the *Aves* and human last common ancestor, some of these similarities may well be analogous, that is, the result of convergent function, rather than homologous, that is, the result of shared ancestry. It remains to be seen whether this remains a sensible conclusion given ongoing research uncovering so-called "deep homologies" among distant ancestors; however, the patchy existence of vocal learning, imitation, and production abilities in both birds and the (far sparser set of) primates suggests that many birdsong-speech commonalities fall more into the class of convergent evolution, while similarities in brain regions, categorical and prosodic perception, and the like may prove to be homologous. While birdsong syntax is more complex than simple bigrams, unlike human language syntax, it does not appear to go beyond languages describable by a narrow class of easily learnable finite-state transition networks. More broadly still, birdsong lacks nearly all the chief attributes of human language: it does not match the syntactic complexity of human language, without the multiple-label, and often asymmetric containment relationships of human language; it does not consist of phrases grounded on the conceptual atoms we call words; and, without words, it does not possess a compositional semantics driven by a full-fledged combinatorial syntax. Nevertheless, as far as a model for human speech acquisition and production, birdsong remains a useful model for the analyzing the evolution of a still-complex interplay between brain and behavior.

ACKNOWLEDGMENTS

Johan J. Bolhuis is funded by Netherlands Organization for Scientific Research (NWO) grants (ALW Open Competition and GW Horizon Programme). Publication of this paper was supported by an NWO Incentive Fund Open Access grant.

REFERENCES

Abe, K., and Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* 14, 1067–1074.

Angluin, D. (1982). Inference of reversible languages. *J. Assoc. Comput. Mach.* 29, 741–765.

Aristotle. (1984). *Historia Animalium*, trans. J. Peck. Cambridge, MA: Harvard University Press [Loeb Classical Library].

Arregui, A., Clifton, C., Frazier, L., and Moulton, K. (2006). Processing elided verb phrases with flawed antecedents. *J. Mem. Lang.* 55, 232–246.

Bartlett, M., and Kazakov, D. (2005). The origins of syntax: from navigation to language. *Conn. Sci.* 17, 271–288.

Barton, E., Ristad, E., and Berwick, R. (1987). *Computational Complexity and Natural Language*. Cambridge, MA: MIT Press.

Beckers, G. J. L., Bolhuis, J. J., Okanoyoa, K., and Berwick, R. (2012). Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* 23, 139–145.

Berwick, R. (1982). *Locality Principles and the Acquisition of Syntactic Knowledge*. Cambridge: MIT Doctoral Dissertation, MIT.

Berwick, R. (1985). *Locality Principles and the Acquisition of Syntactic Knowledge*. Cambridge: MIT Press.

Berwick, R. (2011). "All you need is merge: biology, computation, and language from the bottom-up," in *The Biolinguistic Enterprise*, eds A. M. Di Sciullo and C. Boecxk (Oxford: Oxford University Press), 706–825.

Berwick, R., Beckers, G. J. L., Okanoyoa, K., and Bolhuis, J. J. (2011a). Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* 15, 113–121.

Berwick, R., and Chomsky, N. (2011). "Biolinguistics: the current state of its development," in *The Biolinguistic Enterprise*, eds A. M. Di Sciullo and C. Boeckx (Oxford: Oxford University Press).

Berwick, R., Pietroski, P., Yankama, B., and Chomsky, N. (2011b). Poverty of the stimulus revisited. *Cogn. Sci.* 35, 1207–1242.

Berwick, R., and Pilato, S. (1987). Learning syntax by automata induction. *J. Mach. Learn. Res.* 2, 9–38.

Berwick, R., and Weinberg, A. (1985). *The Grammatical Basis of Linguistic Performance*. Cambridge, MA: MIT Press.

Bever, T. (1970). "The cognitive basis of linguistic structures," in *Cognition and the Development of Language*, ed. J. R. Hayes (New York: Wiley), 279–362.

Bijeljac-Babic, R., Bertoncini, J., and Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Dev. Psychol.* 29, 711–721.

Bloomfield, L. (1933). *Language*. New York: Holt.

Bolhuis, J. J., and Eda-Fujiwara, H. (2003). Bird brains and songs: neural mechanisms of birdsong perception and memory. *Anim. Biol.* 53, 129–145.

Bolhuis, J. J., and Eda-Fujiwara, H. (2010). Birdsong and the brain: the syntax of memory. *Neuroreport* 21, 395–398.

Bolhuis, J. J., and Gahr, M. (2006). Neural mechanisms of birdsong memory. *Nat. Rev. Neurosci.* 7, 347–357.

Bolhuis, J. J., Okanoya, K., and Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* 11, 747–759.

Bolker, J., and Raff, R. (1996). Developmental genetics and traditional homology. *Bioessays* 18, 489–494.

Brainard, M. S., and Doupe, A. J. (2000). Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* 404, 762–766.

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

Brauer, J., Anwander, A., and Friederici, A. (2011). Neuroanatomical prerequisites for language functions in the maturing brain. *Cereb. Cortex* 21, 459–466.

Chater, N., and Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cogn. Sci.* 34, 1131–1157.

Chomsky, N. (1955). *The Logical Structure of Linguistic Theory*. Chicago: University of Chicago Press.

Chomsky, N. (1956). Three models for the description of language. *IEEE Transactions on Information Theory* 2, 113–124.

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

Chomsky, N. (1963). "Formal properties of grammar," in *Handbook of Mathematical Psychology*, eds D. Luce, R. Bush, and E. Galanter (New York: Wiley), 323–418.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

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

Chomsky, N., and Miller, G. (1963). "Finitary models of language users," in *Handbook of Mathematical Psychology*, eds R. Luce, R. Bush, and E. Galanter (New York: Wiley), 419–491.

Corballis, M. C. (2007). Recursion, language and starlings. *Cogn. Sci.* 31, 697–704.

Crain, S., and Nakayama, M. (1987). Structure dependence in question formation. *Language* 62, 522–543.

Culicover, P., and Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.

De Marcken, C. (1995). "Lexical heads, phrase structure and the induction of grammar," in *Proceedings of the Third Workshop on Very Large Corpora*, eds D. Yarowsky and K. Church (Cambridge: Association for Computational Linguistics), 14–26.

De Marcken, C. (1996). *Unsupervised Language Acquisition*. Ph.D. dissertation, MIT, Cambridge.

Dehaene, S. (1997). *The Number Sense*. Oxford: Oxford University Press.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Mériaux, S., Roche, A., Sigman, M., and Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14240–14245.

Di Sciullo, A. M. (2003). *Asymmetry in Grammar: Syntax and Semantics*. New York: John Benjamins.

Doupe, A. J., Perkel, D. J., Reiner, A., and Stern, E. A. (2005). Birdbrains could teach basal ganglia research a new song. *Trends Neurosci.* 28, 353–363.

Dulai, K., von Dornum, M., Mollon, J. D., and Hunt, D. M. (1999). The evolution of trichromatic colour vision by opsin gene duplication. *Genome Res.* 9, 629–638.

Elemans, C. P. H., Mead, A. F., Jakobson, L., and Ratcliffe, J. M. (2011). Superfast muscles set maximum call rate in echolocating bats. *Science* 333, 1885–1888.

Fee, M., Kozhevnikov, A., and Hahnloser, R. H. (2004). Neural mechanisms of vocal sequence generation in the songbird. *Ann. N. Y. Acad. Sci.* 1016, 153–170.

Feher, O., Wang, H., Saar, S., Mitra, P. P., and Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature* 459, 564–568.

Fisher, S. E., and Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends Genet.* 25, 166–177.

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., and Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nat. Genet.* 18, 168–170.

Fitch, W. T. (2005). The evolution of language: a comparative review. *Biol. Philos.* 20, 193–203.

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215.

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.

Fitch, W. T. (2011). "'Deep homology' in the biology and evolution of Language," in *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, eds A. M. Di Sciullo and C. Boeckx (Oxford: Oxford University Press), 135–166.

Fitch, W. T., and Hauser, M. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.

Fodor, J. A. (1996). The pet fish and the red herring: why concepts aren't prototypes, *Cognition* 58, 243–276.

Fong, S., and Di Sciullo, A.-M. (2005). "Morpho-syntax parsing," in *UG and External Systems*, eds A.M. Di Sciullo and R. Delmonte (Amsterdam: John Benjamins), 247–268.

Friederici, A., Bahlmann, J., Friederich, R., and Makuuchi, M. (2011). The neural basis of recursion and complex syntactic hierarchy. *Biolinguistics* 5, 87–104.

Friederici, A. D., and Brauer, J. (2009). "Syntactic complexity in the brain," in *Syntactic Complexity: Diachrony, Acquisition, Neurocognition, Evolution*, eds T. Givon and M. Shibatani (Amsterdam: John Benjamins), 491–506.

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 recognition in European starlings (*Sturnus vulgaris*). *Anim. Behav.* 56, 579–594.

Gobes, S. M. H., and Bolhuis, J. J. (2007). Bird song memory: a neural dissociation between song recognition and production. *Curr. Biol.* 17, 789–793.

Gobes, S. M. H., Zandbergen, M. A., and Bolhuis, J. J. (2010). Memory in the making: localized brain activation related to song learning in young songbirds. *Proc. R. Soc. Lond. B Biol. Sci. B* 277, 3343–3351.

Gold, E. (1978). Complexity of automaton identification from given data. *Inform. Contr.* 37, 302–320.

Goldberg, A. (2006). *Constructions at Work*. Oxford: Oxford University Press.

Haesler, S., Wada, K., Nshdejan, A., Edward, E., Morrisey, E., Lints, T., Jarvis, E. D., and Scharff, C. (2004). FoxP2 expression in avian vocal learners and non-learners. *J. Neurosci.* 24, 3164–3175.

Halder, G., Callaerts, P., and Gehring, W. J. (1995). New perspectives on eye evolution. *Curr. Opin. Genet. Dev.* 5, 602–609.

Halle, M., and Chomsky, N. (1968). *The Sound Patterns of English*. New York: Harcourt-Brace.

Halle, M., and Idsardi, W. J. (1995). "General properties of stress and metrical structure," in *A Handbook of Phonological Theory*, ed. J. Goldsmith (Oxford: Blackwell), 403–443.

Heinz, J. (2010). Learning long-distance phonotactics. *Linguistic Inquiry* 41, 623–661.

Heinz, J., and Idsardi, W. (2011). Sentence and sound complexity. *Science* 333, 295–297.

Hopcroft, J., and Ullman, J. (1979). *Introduction to Automata Theory, Languages, and Computation*. Reading, MA: Addison-Wesley.

Horning, J. (1969). *A Study of Grammatical Inference*. Ph.D. dissertation, Stanford University, Stanford.

Hsu, A. S., Chater, N., and Vitanyi, P. M. B. (2011). The probabilistic analysis of language acquisition: theoretical, computational, and experimental analysis. *Cognition* 120, 380–390.

Hurford, J. R. (2011). *The Origins of Grammar*. Oxford: Oxford University Press.

Huybregts, M. A. C. (1984). "The weak adequacy of context-free phrase structure grammar," in *Van Periferie Naar Kern*, eds G. J. de Haan, M. Trommelen, and W. Zonneveld (Dordrecht, Foris), 81–99.

Ikebuchi, M., and Okanoya, K. (2000). Limited memory for conspecific songs in a non-territorial songbird. *Neuroreport* 11, 3915–3919.

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

Jackendoff, R. (1977). *X-Bar Syntax: A Theory of Phrase Structure*. Cambridge, MA: MIT Press.

Jackendoff, R. (2010). "Your theory of language evolution depends on your theory of language," in *The Evolution of Human Languages: Biolinguistic Perspectives*, eds R. Larson, V. Deprez, and H. Yamakido (Cambridge: Cambridge University Press), 63–72.

Jarvis, E. D. (2007). Neural systems for vocal learning in birds and humans: a synopsis. *J. Ornithol.* 148, 35–54.

Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D. J., Shimizu, T., Striedter, G., Wild, J. M., Ball, G. F., Dugas-Ford, J., Durand, S. E., Hough, G. E., Husband, S., Kubikova, L., Lee, D. W., Mello, C. V., Powers, A., Siang, C., Smulders, T. V., Wada, K., White, S. A., Yamamoto, K., Yu, J., Reiner, A., Butler, A. B., and Avian Brain Nomenclature Consortium. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nat. Rev. Neurosci.* 6, 151–159.

Jarvis, E. D., and Mello, C. V. (2000). Molecular mapping of brain areas involved in parrot vocal communication. *J. Comp. Neurol.* 419, 1–31.

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

Kakishita, Y., Sasahara, K., Nishino, T., Takahasi, M., and Okanoya, K. (2009). Ethological data mining: an automata-based approach to extract behavioural units and rules. *Data Min. Knowl. Discov.* 18, 446–471.

Katahira, K., Suzuki, K., Okanoya, K., and Okada, M. (2011). Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. *PLoS ONE* 6, e24516. doi:10.1371/journal.pone.0024516

Kayne, R. S. (1994). *The Antisymmetry of Syntax*. Cambridge: MIT Press.

Kleene, S. (1953). *Introduction to Metamathematics*. Amsterdam: North-Holland.

Kleene, S. C. (1956). "Representation of events in nerve nets and finite automata," in *Automata Studies*, eds C. Shannon and W. Ashby (Princeton: Princeton University Press), 3–42.

Lambek, J. (1958). The mathematics of sentence structure. *Am. Math. Mon.* 65, 154–170.

Langendoen, T. (1975). Finite-state parsing of phrase-structure languages and the status of readjustment rules in the lexicon. *Linguist. Inq.* 6, 533–554.

Lasnik, H. (2000). *Syntactic Structures Revisited*. Cambridge, MA: MIT Press.

Laurin, M., and Reisz, R. (1995). A re-evaluation of early amniote phylogeny. *Zool. J. Linn. Soc.* 113, 165–223.

Marantz, A. (1982). Re-reduplication. *Linguist. Inq.*13, 435–482.

Marler, P. (1998) "Animal communication and human language," in *The Origin and Diversification of Human Language*, eds N. G. Jablonski and L. E. Aiello (San Francisco, CA: California Academy of Sciences), 1–19.

Mauner, G. A., Tanenhaus, M. K., and Carlson, G. N. (1995). A note on parallelism effects on processing verb phrase anaphors. *Lang. Cogn. Process.* 10, 1–12.

McNaughton, R., and Yamada, H. (1960). Regular expressions and state graphs for automata. *IEEE Trans. Electr. Comput.* 9, 39–47.

Miller, G., and Chomsky, N. (1963). "Finitary models of language users," in *Handbook of Mathematical Psychology*, Vol. 2, eds D. Luce, R. Bush, and E. Eugene Galanter (New York: Wiley), 419–491.

Miller, G. A., and Isard, S. (1964). Free recall of self-embedded English sentences. *Inform. Contr.* 7, 292–303.

Minsky, M. (1967). *Computation: Finite and Infinite Machines*. Englewood Cliffs, NJ: Prentice-Hall.

Mooney, R. (2009). Neural mechanisms for learned birdsong. *Learn. Mem.* 16, 655–669.

Moorman, S., Mello, C. V., and Bolhuis, J. J. (2011). From songs to synapses: molecular mechanisms of birdsong memory. *Bioessays* 33, 377–385.

Moro, A. (2000). *Dynamic Antisymmetry*. Cambridge: MIT Press.

Moro, A. (2008). *The Boundaries of Babel: The Brain and the Enigma of Impossible Languages*. Cambridge, MA: MIT Press.

Moro, A. (2011). A closer look at the turtle's eyes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2177–2178.

Musso, M., Moro, A., Glauche. V., Rijntjes, M., Reichenbach, J., Büchel, C., and Weiller, C. (2003). Broca's area and the language instinct. *Nat. Neuro.* 6, 774–781.

Niyogi, P. (2006). *The Computational and Evolutionary Nature of Language*. Cambridge, MA: MIT Press.

Okanoya, K. (2007). Language evolution and an emergent property. *Curr. Opin. Neurobiol.* 17, 271–276.

Oldfield, J., Williams, K., and Wiseman, N. (1987). Content-addressable memories for storing and processing recursively subdivided images and tress. *Electron. Lett.* 23, 262–263.

Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2522–2527.

Perfors, A., Reiger, C., and Tenenbaum, J. (2010). The learnability of abstract syntactic principles. *Cognition* 3, 306–338.

Perruchet, P., and Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans? *Psychon. Bull. Rev.* 12, 207–313.

Petersson, K. M., Folia, V., and Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain Lang.* 120, 83–95.

Petitto, L. A. (2005). "How the brain 2732 begets language," in *The Cambridge Companion to Chomsky*, ed. J. McGilvray (Cambridge: Cambridge University Press), 84–101.

Petitto, L. A., Holowka, S., Sergio, L., Levy, B., and Ostry, D. (2004). Baby hands that move to the rhythm of language: hearing babies acquiring sign languages babble silently on the hands. *Cognition* 9, 43–73.

Petri, J., and Scharff, C. (2011). Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Phil. Trans. R. Soc. B* 1574, 2124–2140.

Pulvermüller, F. (1993). "On conecting syntax and the brain," in *Brain Theory – Spatio-Temporal Aspects of Brain Function*, ed. Aertsen (New York: Elsevier), 131–145.

Pulvermüller, F., and Knoblauch, A. (2009). Discrete combinatorial circuits emerging in neural networks: a mechanism for rules of grammar in the human brain? *Neural. Netw.* 22, 161–172.

Rabin, M., and Scott, D. (1959). Finite automata and their decision problems. *IBM J. Res. Dev.* 3, 114.

Radford, A. (1997). *Syntactic Theory and the Structure of English: A Minimalist Approach.* Cambridge: Cambridge University Press.

Riebel, K., and Slater, P. J. B. (2003). Temporal variation in male chaffinch song depends on the singer and the song type. *Behaviour* 140, 269–288.

Rogers, J., and Pullum, G. K. (2011). Aural pattern recognition experiments and the subregular hierarchy. *J. Log. Lang. Inform.* 20, 329–342.

Saffran, J., Aslin, R., and Newport, E. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Sag, I. A., Wasow, T., and Bender, E. (2003). *Syntactic Theory: A Formal Introduction*, 2nd Edn. Stanford: CSLI Publications.

Sasahara, K., Kakishita, Y., Nishino, T., Takahasi, M., and Okanoya, K. (2006). "A reversible automata approach to modeling birdsongs," in *Proceedings of 15th International Conference on Computing (CIC2006)* (New York: IEEE Computer Society Press), 80–85.

Scharff, C., and Petri, J. (2011). Evo-devo, deep homology and FoxP2: implications implications for the evolution of speech and language. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 2124–2140.

Shettleworth, S. J. (2010). *Cognition, Evolution, and Behavior.* Oxford: Oxford University Press.

Shubin, N., Tabin, C., and Carroll, S. (1997). Fossils, genes and the evolution of animal limbs. *Nature* 388, 639–648.

Shukla, M., White, K. S., and Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6038–6043.

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

Solomonoff, R. (1964). A formal theory of inductive inference. *Inform. Contr.* 7, 1–22.

Stolcke, A. (1994). *Bayesian Learning of Probabilistic Learning Models.* Ph.D. dissertation, Stanford University, Stanford.

Steedman, M. (2000). *The Syntactic Process.* Cambridge, MA: MIT Press.

Stevens, K. (2000). *Acoustic Phonetics.* Cambridge, MA: MIT Press.

Suge, R., and Okanoya, K. (2010). Perceptual chunking in the self-produced songs of Bengalese Finches (*Lonchura striata* var. *domestica*). *Anim. Cogn.* 13, 515–523.

Suh, A., Paus, M., Kiefmann, M., Churakov, G., Franke, F., Brosius, J., Kriegs, J., and Schmitz, J. (2011). Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat. Commun.* 2, 1–7.

Todt, D., and Hultsch, H. (1996). "Acquisition and performance of repertoires: ways of coping with diversity and versatility," in *Ecology and Evolution of Communication*, eds D. E. Kroodsma, and E. H. Miller (Ithaca, NY; Cornell University Press), 79–96.

Uddén, J., Ingvar, M., Hagoort, P., and Petersson, K. (2011). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: investigating the push-down stack model. *Cogn. Sci.* (in press).

Vernes, S. C., Oliver, P. L., Spiteri, E., Lockstone, H. E., Puliyadi, R., Taylor, J. M., Ho, J., Mombereau, C., Brewer, A., Lowy, E., Nicod, J., Groszer, M., Baban, D., Sahgal, N., Cazier, J.-B., Ragoussis, J., Davies, K. E., Geschwind, D. H., and Fisher, S. E. (2011). Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet.* 7, e1002145. doi:10.1371/journal.pgen.1002145

Webb, D., and Zhang, J. (2005). FoxP2 in song-learning birds and vocal-learning mammals. *J. Hered.* 96, 212–216.

Weir, D. (1988). *Characterizing Mildly-Context Sensitive Grammar Formalisms.* Ph.D. dissertation, University of Pennsylvania, Philadelphia.

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

Woods, W. (1970). Transition network grammars for natural language analysis. *Commun. ACM* 13, 591–606.

Yip, M. (2006). The search for phonology in other species. *Trends Cogn. Sci.* 10, 442–446.

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: 01 December 2011; accepted: 20 March 2012; published online: 13 April 2012.

Citation: Berwick RC, Beckers GJL, Okanoya K and Bolhuis JJ (2012) A bird's eye view of human language evolution. Front. Evol. Neurosci. ***4****:5. doi: 10.3389/fnevo.2012.00005*

frontiers in
EVOLUTIONARY NEUROSCIENCE

REVIEW ARTICLE
published: 09 April 2012
doi: 10.3389/fnevo.2012.00006

Targets for a comparative neurobiology of language

Justin T. Kiggins[1], Jordan A. Comins[2] 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).

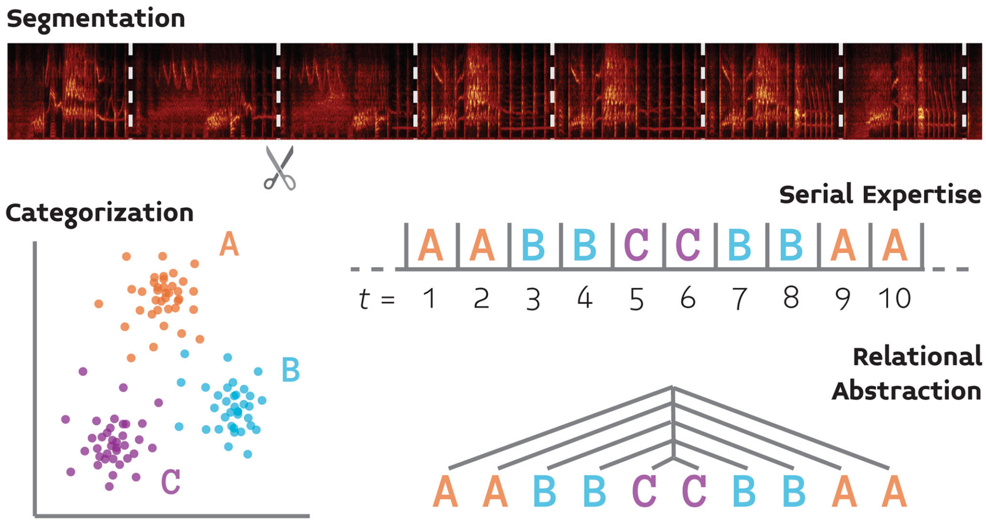


FIGURE 1 | A schematic illustration of four processes (segmentation, categorization, serial processing, and relational abstraction) which support auditory cognition and for which we propose that comparative behavioral and neurobiological experiments will yield a fuller understanding of language perception.

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 infanthood (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 pseudowords. 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 make 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 actively 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 an many other similar single neuron studies (see Miller, 2007 for recent review) may direct 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 task 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, F., 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 birdsong 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.

Iversen, 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.

Liberman, 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. R., 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.

Takahasi, 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*

frontiers in
EVOLUTIONARY NEUROSCIENCE

REVIEW ARTICLE
published: 16 August 2012
doi: 10.3389/fnevo.2012.00012

Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates

Christopher I. Petkov[1,2]* and Erich D. Jarvis[3]*

[1] Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK
[2] Centre for Behavior and Evolution, Newcastle University, Newcastle upon Tyne, UK
[3] Department of Neurobiology, Howard Hughes Medical Institute, Duke University, Durham, NC, USA

Edited by:
Angela D. Friederici, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:
Josef P. Rauschecker, Georgetown University School of Medicine, USA
Kazuo Okanoya, The University of Tokyo, Japan

***Correspondence:**
Christopher I. Petkov, Institute of Neuroscience, Newcastle University Medical School, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK.
e-mail: chris.petkov@ncl.ac.uk

Erich D. Jarvis, Department of Neurobiology, Howard Hughes Medical Institute, Box 3209, Duke University Medical Center, Durham, NC 27710, USA.
e-mail: jarvis@neuro.duke.edu

Vocal learners such as humans and songbirds can learn to produce elaborate patterns of structurally organized vocalizations, whereas many other vertebrates such as non-human primates and most other bird groups either cannot or do so to a very limited degree. To explain the similarities among humans and vocal-learning birds and the differences with other species, various theories have been proposed. One set of theories are motor theories, which underscore the role of the motor system as an evolutionary substrate for vocal production learning. For instance, the motor theory of speech and song perception proposes enhanced auditory perceptual learning of speech in humans and song in birds, which suggests a considerable level of neurobiological specialization. Another, a motor theory of vocal learning origin, proposes that the brain pathways that control the learning and production of song and speech were derived from adjacent motor brain pathways. Another set of theories are cognitive theories, which address the interface between cognition and the auditory-vocal domains to support language learning in humans. Here we critically review the behavioral and neurobiological evidence for parallels and differences between the so-called vocal learners and vocal non-learners in the context of motor and cognitive theories. In doing so, we note that behaviorally vocal-production learning abilities are more distributed than categorical, as are the auditory-learning abilities of animals. We propose testable hypotheses on the extent of the specializations and cross-species correspondences suggested by motor and cognitive theories. We believe that determining how spoken language evolved is likely to become clearer with concerted efforts in testing comparative data from many non-human animal species.

Keywords: evolution, humans, monkeys, avian, vertebrates, communication, speech, neurobiology

INTRODUCTION

Charles Darwin's theory on descent with modification as it applies to man (Darwin, 1871) had for many years been used to underscore the importance of non-human primates for unraveling the origins and neuronal precursors of spoken language (e.g., Hewes, 1973). Yet, in part because of the apparent lack of vocal learning or syntactic-like abilities in non-human primates, different camps have focused on either the differences between human and non-human primates or their similarities. This has resulted in many contentious debates on language evolution with regards to non-human primates (for reviews: Pinker, 1994; Hauser et al., 2002; Pinker and Jackendoff, 2005). Adding further complexity for understanding spoken language origins, recently the research focus has shifted towards species more distantly related to humans, such as certain groups of songbirds. This is in part because songbirds like humans and a few other species exhibit vocal learning and have what has been broadly classified as "syntactic-like" song production (Doupe and Kuhl, 1999; Jarvis, 2004; Bolhuis et al., 2010). A summary of a consortium on the origins of human language syntax and its biological foundations encapsulates some of the current thinking:

> Another area of agreement might seem surprising in light of many current "primate-centric" studies of language evolution (Burling, 2006; Hurford, 2007). Most participants felt that there were no true precursors of syntax to be found among our nearest relatives. For anything like a syntactic precursor one had to go as far afield as songbirds ....
>
> (Bickerton and Szathmary, 2009)

Likewise, in a thought provoking essay, Bolhuis and Wynne (2009) questioned to what extent evolutionary theory can help us to understand cognitive brain mechanisms in living animals. Their perspective was illustrated by a cartoon depicting a scientist with the great hope of teaching a monkey to say "apple," but realizing that the monkey is the classroom dunce when the parrot vocally identifies the apple variety as "golden delicious." Darwin, however, would have likely filled the classroom with as many different animals as possible. In any case, the authors' conclusions are appropriately nuanced and seem to favor a broader comparative approach: "there is no a priori reason to assume that convergence will be more important than common descent or vice versa" (Bolhuis and Wynne, 2009).

We, as researchers that have studied non-human primates and birds, argue that the path toward understanding the origins of spoken language cannot be based on focusing on a few select species with or without communication abilities that are either thought to be most comparable to humans, or to reflect physiology most comparable to humans. Any "one animal centric" approach will only limit our capacity to unravel the evolutionary bases of spoken language. If for no other reason, without "other" species as points of reference, it would not be clear what is special about human communication. Moreover, a focus on certain species restrains the development of different animal model systems with distinct advantages for understanding the neurobiological mechanisms of human language-related processes, which is important for advancing treatment options for communication and language disorders. Thus, to better understand the origins of human spoken language we rely on a broad comparative approach that takes advantage of information obtained across animal taxa, letting each animal have their "say" on the question of language evolution. We are aware that to do so can only be achieved by additional comparative work that will require energy and investment, combined with efforts to stay objective, as best as we can, regarding the cross-species similarities and differences.

In an effort to invigorate a broader perspective on spoken language origins, in this paper we overview the parallels and differences in the behavioral and neurobiological data of vocal learners (e.g., humans and songbirds) and those animals often identified as "vocal non-learners." We ask how strong is the evidence for categorical distinctions between vocal learners and vocal non-learners? We note that vocal non-learners are often classified as such based on a lack of experimental evidence, but that when the animals are tested, there is often more variation in vocal learning abilities than might have been expected (Janik and Slater, 2000; Snowdon, 2009; Arriaga et al., in press). Moreover, since vocal learning depends on auditory learning, and auditory learning abilities are broadly conserved in the animal kingdom, we ask how this trait dependency could have influenced the evolution and mechanisms of vocal learning. Then, based on a modified perspective of the literature we reconsider some of the motor and other theories that have been proposed for humans, birds and other animals. We conclude by generating testable hypothesis, including for: (1) better understanding variability in the vocal behavior and neurobiology of vertebrates that are often classified as vocal non-learners; and (2) the possible capabilities of, for example, non-human primates as limited vocal learners but considerable auditory learners, to learn the structure of auditory sequences, and whether this might tap into an ancestral "proto-syntactic" brain network that evolved in humans to support syntactic learning.

VOCAL PRODUCTION LEARNING AND AUDITORY LEARNING: HOW ARE THESE BEHAVIORAL PHENOTYPES DISTRIBUTED?

Behavioral data demonstrating that an animal can learn to produce novel vocalizations is often used to classify different species as either vocal learners or vocal non-learners (Nottebohm, 1976; Janik and Slater, 1997; Jarvis, 2004). However, once some animals within a taxonomic group are characterized as vocal learners, we cannot assume that all animals of that group have vocal production learning abilities to the same degree. For instance, different song learning birds have different levels of complexity in their song production, and humans (including infants) can be regarded as exceptional vocal learners (i.e., *high-end* of *vocal learners*, see **Figure 1**). Among passerine songbirds, some species learn to produce only one song that was learned early in life, while others can learn many songs with some level of continuous learning throughout adulthood (Catchpole and Slater, 1995; Okanoya, 2004). For example, songbirds such as zebra finches tend to learn one song type as juveniles. Such songs often have strictly-linear transitions that step through the different song syllables in a motif from beginning to end (Honda and Okanoya, 1999). On the other hand, the songs of mockingbirds, nightingales and humpback whales show considerably greater variability. Some of these song elaborations show repetitions of particular elements within a range of legal repetitions and can include forward or backward branching relationships in how the animals transition between the different elements of their song, as well as non-adjacent relationships between distant song elements. Such "syntactic-like" structure in songbirds has drawn the interest of linguists and cognitive neuroscientists (e.g., Bickerton and Szathmary, 2009; Berwick et al., 2011; Hurford, 2012).

Moreover, not all vocal learners are known to be able to imitate the vocalizations of other species, called vocal mimicry. Yet, one of the initial tests that some have used as evidence to conclude that non-human primates are not vocal learners was the difficulty that chimpanzees have in imitating human speech (e.g., Rumbaugh, 1977; Shettleworth, 2010). Some animals such as corvid songbirds (e.g., crows, jays and magpies) and African Grey and Amazon parrots are exceptional imitators that can imitate human speech (e.g., Kroodsma, 1976; Pepperberg, 2010). The birds at this end of the spectrum are fairly easy to identify since they often imitate without training or an obvious reward. Moore (2004) notes that, "Moore's (1992) parrot, for example, repeatedly mimicked a swear word that it had heard only once, through a closed door."

Another important issue is that the distinction between vocal learning and non-learning captures only the more apparent differences in vocal production behavior. Song learning in birds and speech learning in humans takes at least two phases: auditory learning and sensory-motor vocal learning (Doupe and Kuhl, 1999). Only the latter is evident in vocal production behavior. For example, many of us are familiar with the situation where as we learn a second language we progress through stages where our ability to understand the language outstrips our ability to produce it. Thereby, in discussing how the human brain has specialized to support spoken language, not only do we need to consider the cross-species variability of specific behavioral phenotypes, but we also need to distinguish different behavioral phenotypes, such as sensory (auditory) learning and vocal production learning (Jarvis, 2004).

In this section, we distinguish between vocal production and auditory learning abilities and consider how these two behaviors

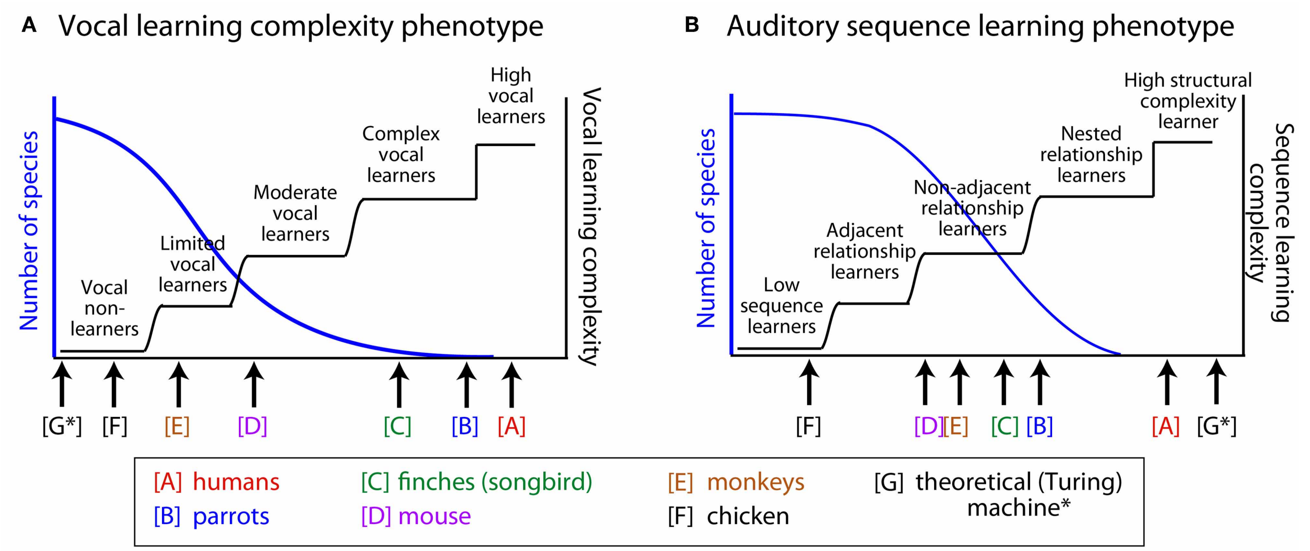


FIGURE 1 | Hypothetical distributions of two behavioral phenotypes: vocal learning and sensory (auditory) sequence learning. We hypothesize that the behavioral phenotypes of vocal learning and auditory learning are distributed along several categories. **(A)** Vocal learning complexity phenotype and **(B)** auditory sequence learning phenotype. The left axis (blue) illustrates the hypothetical distribution of species along the behavioral phenotype dimensions. The right axis (black step functions) illustrates different types of transitions along the hypothesized vocal-learning **(A)** or auditory-learning **(B)** complexity dimensions. See manuscript text for the basis for the relative position of the non-human animals illustrated in this figure, which in some cases is based on limited data. Also see Arriaga and Jarvis (in press) for an initial proposal of this idea. Whether the actual distributions are continuous functions (blue curves), will need to be tested, in relation to the alternatives that there are several categories with gradual transitions or step functions (black curves). Although auditory learning is a prerequisite for vocal learning and there can be a correlation between the two phenotypes (A–B), the two need not be interdependent. A theoretical Turing machine (Turing, 1968) is illustrated [G*], which can outperform humans on memory for digitized auditory input but is not a vocal learner.

might be distributed. Why does the variance in a behavioral phenotype matter? It is important to determine the ways in which, for example, vocal learning is distributed since each possibility carries with it different implications for: (1) how spoken language is likely to have originated; (2) the evolutionary pressures that regulate the presence and absence of a behavioral phenotype; and, (3) whether few or many animals might serve to model certain mechanistic aspects of human speech- and language-related processes.

We note that when the findings of recent studies are examined across species, there seems to be evidence for greater variability in the experience-based ability to modify vocalizations. This variability is greater than would be expected in animals often assumed to be vocal non-learners (e.g., Saranathan et al., 2007; Snowdon, 2009; Briefer and McElligott, 2011; Arriaga et al., in press). Our review of the literature suggests that the currently known vocal learners could be more accurately described as *complex-vocal leaners* (humans potentially different among these as *high vocal learners*), and other species as *moderate-vocal learners*, *limited-vocal learners*, or obligate *vocal non-learners*. Thereby, the empirical evidence does not fit well with a binary categorical distinction between vocal learning and non-learning. Instead, the evidence better fits with the notion of a recently proposed vocal-learning continuum hypothesis (Arriaga and Jarvis, in press). We conclude by considering how approaches in linguistics can be useful for unraveling the complexity of animals' vocal production and/or auditory learning capabilities.

VARIABILITY IN THE VOCAL LEARNING PHENOTYPE

We begin with the useful designations of vocal learning made by Janik and Slater (1997), but with our modifications of their three categories as: production vocal learning; usage vocal learning; and, auditory learning (instead of comprehension learning).

Production vocal learning is often defined as the ability of an animal to produce novel vocalizations. There are various ways in which this can be tested. These include: (1) cross-fostering experiments, such as when an infant can learn the vocalizations of surrogate parents of the same or other species; (2) experiments evaluating changes in vocalizations in response to different types of competing sounds; (3) social isolation studies where the animal does not have access to a model to imitate; and (4) deafening experiments where the animal cannot hear others or itself. The strongest evidence that vocal production learning has occurred is demonstrating that an animal can imitate vocalizations of other species or sounds. Vocal imitation depends upon the animal being able to hear and to have a model to imitate. As such, deaf *vocal learners* usually show acoustically degraded, species non-typical sounds, whereas deaf *vocal non-learners* typically show vocal production behavior that does not differ from wild type animals. Sometimes an argument is made that the effects of social isolation or deafening could be due to unspecified impacts on behavior. This becomes less of an issue if it is shown that the behavioral impact is specific to the vocal learning modality.

Another careful distinction that needs to be made is defining the anatomical source of the "vocalizations" that are found to be learned. Sounds generated by the vocal organ (larynx in mammals; syrinx in birds) are in the strictest sense, vocalizations, whereas those generated by lips, teeth, and tongue are non-voiced, oro-facially generated or modified sounds. The neural mechanisms and the degree of control of the different musculature for generating voiced or unvoiced sounds could differ. Whatever the approach, to substantiate that an animal is capable of production vocal learning one needs to document a convincing experience-dependent change in vocal production behavior, with consideration of the source of the sounds.

Usage vocal learning is when an animal learns to use acoustically innate or already learned vocalizations in a new context. Although usage vocal learning involves the learning of the contexts in which to vocalize, it is not production vocal learning because it does not require modification of the acoustic structure of a vocalization to create one that the animal did not have in its repertoire. Common examples of usage vocal learning are the predator alarm calls of vervet monkeys and chickens, where juveniles learn through social experience the context within which to generate the innate call or the appropriate behavioral response (e.g., descend from trees, tilt head, etc.) upon hearing the call from a conspecific (Seyfarth et al., 1980; Evans et al., 1993; Snowdon, 2009).

Auditory learning is when an animal learns to perceive something novel or behaviorally react to sounds differently as a result of experience. For example, a dog learns how to associate the human sounds "sit" or "sientese" with the act of sitting, but it does not learn how to produce or vocally use these sounds. Even with this capacity, there might be limits to the complexity of verbal commands that an animal might be able to learn in the auditory modality (Moore, 2004).

Who are the production vocal learners?

All vocal species tested appear to have varying degrees of auditory learning and usage vocal learning (Moore, 2004; Schusterman, 2008), but only relatively few have production vocal learning (Janik and Slater, 1997; Jarvis, 2004). The later include, in addition to humans, three groups of birds (passerine songbirds, parrots and hummingbirds; Marler and Tamura, 1964; Jarvis, 2004; Jarvis et al., 2005; Bolhuis et al., 2010; Pepperberg, 2010), some species of bats (Esser, 1994; Boughman, 1998), and pinnipeds and cetaceans (Noad et al., 2000). For example, in several cases, a harbor seal and dolphin were found to imitate human vocalizations (Lilly, 1965; Ralls et al., 1985). This list has recently expanded to include elephants (Poole et al., 2005), where an African elephant was shown to imitate Asian elephant calls and another elephant to imitate the sounds of passing trucks. In the rest of this paper we will refer to these species as vocal learners, meaning production vocal learners.

Who are the vocal non-learners?

The answer to this question is much less clear. In contrast to the few known vocal learning species, it is commonly thought that most other vertebrates are not capable of vocal imitation: that is, they are not capable of the type of learning that leads to the production of novel communication signals not within their repertoire or to the production of acoustic changes to innate vocalizations. However, most of these animals have not been formally or rigorously tested to determine whether they have more limited capabilities for some aspects of production vocal learning. That is, many vertebrates are usually placed in the "vocal non-learning" category based largely on a lack of evidence rather than evidence for a lack of any vocal learning capabilities. In the last few decades, with the use of refined acoustical analysis tools and hypothesis-driven experimental strategies, evidence appears to be accumulating that some of the, so-called, vocal non-learners have measurable levels of vocal flexibility to change some of the acoustics in their vocalizations, potentially a limited form of vocal learning. We will consider several examples of this from work in non-human primates (Snowdon, 2009), birds (Saranathan et al., 2007), rodents (Arriaga et al., in press), and goats (Briefer and McElligott, 2011).

For non-human primates, the earlier studies are difficult to interpret, in part because of contradictory conclusions. One study reported what seemed like ontogenetic acoustical changes to innate vocalizations in a cross-fostering study of two species of macaques (Masataka and Fujita, 1989). This finding was later challenged by lack of independent replication of the findings and on technical grounds (Owren et al., 1992). For a review see (Egnor and Hauser, 2004). Furthermore, many of the more striking examples of non-human primates having changed their vocalizations have been shown between regionally separated animals, which could be based primarily on genetically regulated differences between populations (Snowdon, 2009). More recently, a number of studies have shown that non-human primates can make some limited modifications to their presumably innate vocalizations via laryngeal control. For instance, there are several examples of developmental maturation of vocalizations toward their adult form that might not be entirely consistent with innate developmental changes in the vocal production apparatus (e.g., vervets: Seyfarth and Cheney, 1986; prosimians: Zimmerman, 1989; squirrel monkeys: Hammerschmidt et al., 2001). Additional evidence that could question the "vocal non-learning" label in non-human primates has been from call convergence studies, where either two marmosets or macaques housed together for a few weeks showed convergence in the pitch and some other acoustic features of innately determined calls (reviewed in: Snowdon, 2009). Also, there is evidence that adult Japanese macaques are able to adjust the fundamental frequency of their innately-specified vocalizations to match playbacks presented to them of different conspecifics (Sugiura, 1998).

It is important to not only consider laryngeal control in the modification of vocalizations since, for example, human speech is modified by labial and oro-facial control. In this regard, some non-human primates seem to show relatively greater control in modifying the acoustics of their vocalizations and/or to produce non-vocal sounds that do not appear to be innate. Chimpanzees can produce novel attention-getting sounds that are modified by labial (lip) vibrations (Hopkins et al., 2007). This includes a "Raspberry" call where the animals pucker the lips and make a vibrating sound (http://www.youtube.com/watch?v=2Oma_2RFTaM). This call can be imitated by naïve individuals in captivity and some of these calls are also seen in populations in the wild (Marshall et al., 1999;

Crockford et al., 2004; Egnor and Hauser, 2004). Consistent with these observations, an orangutan learned to use her lips to copy a novel whistle produced by a human (Wich et al., 2009).

Certainly, relative to song-learning birds, humans and other vocal learners, non-human primates do not fit the stereotyped view of a production vocal learner (Egnor and Hauser, 2004; Snowdon, 2009). Rather, we would interpret the evidence for vocal plasticity and flexibility in some non-human primates as limited-vocal learning, albeit with greater flexibility via non-laryngeal than laryngeal control. But they do not have the considerable levels of laryngeal (mammalian) or syringeal (avian) control as seen in complex vocal learners. We next ask: are there other examples in the animal literature that also do not clearly fit the categorical vocal learning vs. vocal non-learning distinction?

Non-primate examples challenging the vocal learning/non-learning distinction

Regarding the so-called vocal *non-learning* birds, there is an interesting report on a suboscine passerine bird with possible evidence of at least limited-vocal learning (Saranathan et al., 2007). Suboscines are the closest relatives of songbirds, like chimpanzees are to humans. Likewise the first suboscine species studied (Eastern Phoebe—Saynoris genus; Kingbird—Tyrannus genus) were found to: (1) not be vocal learners based on social isolation and deafening experiments; and (2) to not have forebrain song nuclei as seen in their close oscine passerine songbird relatives (Nottebohm, 1980; Kroodsma and Konishi, 1991). However, another species belonging to a separate suboscine lineage, the Three-wattled bellbird (Procnias genus) shows conspicuous ontogenetic and geographical song variation and fairly rapid song change within a population, which the authors argue cannot be explained by genetic differences alone (Saranathan et al., 2007). These results suggest that this suboscine species could be a limited- or even moderate-vocal learner, rather than a vocal non-learner. Determining the presence or absence of song nuclei in their forebrain still needs to be investigated.

Mouse ultrasonic vocalizations have recently been described as having "song" or song-like characteristics (Holy and Guo, 2005). However, this does not necessarily mean that mice are vocal learners, because, for example, in birds songs can either be learned or innate (Kroodsma and Konishi, 1991). Kikusui et al. (2011) conducted cross-fostering experiments with mice and did not find evidence of vocal learning (Kikusui et al., 2011). However, recent work by Arriaga and Jarvis (Arriaga and Jarvis, in press; Arriaga et al., in press) on cross-housed males shows that the animals sing their ultrasonic courtship "song" to females with a different pitch in the presence of other males from different strains. For instance, one mouse will match the pitch of his larger male cage mate in the presence of a female. In addition, deafening showed that the mice require auditory feedback to develop and maintain some of the acoustic properties of their song syllables (Arriaga and Jarvis, in press; Arriaga et al., in press). This seems to be limited vocal learning, because the animals appear to be making acoustic changes to innately specified vocalizations.

As another example, a recent report in an ungulate (goats) shows what we believe to be limited-vocal learning (Briefer and McElligott, 2011). The authors studied the social effects of goat vocal ontogeny and note a number of acoustical differences between kids that were placed in different social groups. Goats in the same groups showed more similarity in vocalization acoustics. Here, again changes seem to occur to innate pre-specified vocalization components, to the point that there are considerable differences in the final modified vocalizations relative to the originals.

These examples in the animal behavior literature suggest a greater variability in vocal flexibility than is often appreciated. Certainly, some animals would likely remain in the "vocal non-learner" category, for which there is considerable evidence for a lack of vocal flexibility under different conditions. Yet, findings such as illustrated by the above examples provide support for the vocal learning continuum hypothesis (Arriaga and Jarvis, in press) and we would suggest that certain birds, non-human primates, mice, and goats could be reclassified as either *limited-vocal learners* or *moderate-vocal learners*, including many currently known vocal learners as *complex-vocal learners* (with humans among these as *high-vocal learners*). This hypothesis is illustrated in **Figure 1A**, where the hierarchically higher the vocal-learning category, the fewer species that are observed in that category. We next consider whether there is any evidence that sensory learning, auditory sequence learning in particular, is similarly distributed across several categories, and if so, how could it have influenced the evolution and mechanisms of vocal learning.

COMPLEXITY IN VOCAL PRODUCTION VS. SENSORY (AUDITORY) LEARNING: EVALUATING SYNTACTIC-LIKE VOCAL PRODUCTION AND HOW ANIMALS LEARN ARTIFICIAL GRAMMARS

It has been argued that a distinction needs to be drawn between production vocal learning and sensory (e.g., auditory) learning (Jarvis, 2004; Petkov and Wilson, 2012). Auditory learning appears to be more broadly distributed in the animal kingdom than production vocal learning, and, although necessary, it is not sufficient for vocal learning. For instance, certain dogs, in which there is no evidence for complex vocal production learning, can be trained by humans to associate the sounds of spoken names of tens to hundreds of objects and to retrieve either the correctly named objects or novel objects (Kaminski et al., 2004). Almost all animals tested in classical conditioning experiments can learn to make simple sound associations with reward or punishment, such as detecting single sounds or discriminating pairs of differing sounds (Moore, 2004). The question we ask is what is the range of auditory learning complexity across species and how might this relate to human syntactic learning capabilities? In this regard, it is useful to look at the interface of linguistic theory and experiments in evolutionary biology, which aim to address the level of "syntactic-like" sequencing complexity in either vocal production or the sensory learning capabilities of different animals.

Human syntactic abilities allow us to both perceive and produce grammatical relations between words or word parts in a sentence, and linguists distinguish between language competence and language performance (Chomsky, 1965). Modern linguistic theory has been applied to characterize not only human syntactic abilities but also the complexity in vocal production or auditory sequence learning capacities in a variety of non-human animals (Okanoya, 2004; Berwick et al., 2011; Hurford, 2012). For instance, the Formal Language Hierarchy (FLH) contains several categories of grammar (rule-based systems), each describing an increasingly powerful computational language (Chomsky, 1957; Berwick et al., 2011; Hurford, 2012). Lower ranked grammars, called Finite-State Grammars (FSG) are computationally weaker systems that can only generate strings of sequences with limited structural complexity. Higher ranked grammars can also generate the simpler forms of structural complexity but are less limited. Human spoken language is said to encompass the later, as it can have elaborate hierarchical structures with many non-adjacent relationships between sequence elements, such as the nesting of phrases within other phrases (Berwick et al., 2011; Hurford, 2012; Jaeger and Rogers, 2012; Petkov and Wilson, 2012). Such abilities are thought to be unique to humans in both production and perception. Some animal behavioral studies have challenged this perspective, but remain highly controversial (for a review: Berwick et al., 2011; Jaeger and Rogers, 2012; Ten Cate and Okanoya, 2012). We argue that, instead of focusing on the threshold of "human unique" capabilities, further efforts are needed to better resolve the different levels of complexity in the FLH where non-human animal capabilities are likely to vary to a greater extent (see: Hurford, 2012; Jaeger and Rogers, 2012; Petkov and Wilson, 2012). Combined with further comparative testing, this approach could provide novel insights on the relationship between animal sequence learning capabilities either for perception or production and human syntactic capabilities.

Structural complexity of animal vocal production

As complex vocal learning groups, songbirds and whales are known to naturally produce sequences of their songs with syntactic-like organization, but the structure of their songs do not seem to be more elaborate than sequences that can be generated by FSGs (or "regular grammars") (Okanoya, 2004; Bolhuis et al., 2010; Berwick et al., 2011). In other words, unlike humans, non-human animals do not seem to show deeper hierarchical relationships, such as the nesting of song phrases within others. Further, humans can change the meaning of expressions by changing the syntactic organization of the units, called "compositional syntax" (Tallerman, 2011; Hurford, 2012). But the songs of non-human animals have so far been only characterized as "phonological syntax," since the way that the units are structured are thought not to generate new meanings (Marler, 1970, 2000; Berwick et al., 2011). It remains possible that further experiments with many more species could obtain data to challenge these interpretations of the animal behavioral literature.

As for vocal non-learners or limited-vocal learners, the natural syntactic-like vocal production abilities of non-human primates and many other vertebrates seem to be considerably more limited than those of complex-vocal learners. For example, some species of guenons (Old World monkeys) appear to combine pairs of calls into different context-specific call sequences (Ouattara et al., 2009). Other guenon species use combinations of two alarm calls to elicit group movement in the wild that does not seem to be instigated by the individual calls themselves or by other types of call sequences (Arnold and Zuberbuhler, 2006). Whether other non-human primate species can use and produce combinations of call pairs is currently unknown. It has been suggested that gibbon "song-like" vocalizations contain a different organization of vocalizations when predators are present (Clarke et al., 2006). However, it is not clear whether the information bearing parameters of gibbon songs lie in the proportion of particular song elements and/or the structure of how the elements are organized. Chimpanzees are able to learn to manually combine several learned visual symbols to "sign" with humans (Rumbaugh, 1977), but their ability to do so with vocalizations is considerably more limited and in all cases these abilities require extensive training (Shettleworth, 2010). Therefore, the current impression is that the combinatorial vocal production capabilities of non-human primates are limited to combinations of one to two vocalizations.

Artificial-Grammar Learning and animal sequence learning capabilities

Just as vocal production capabilities seem to vary in complexity across the animal kingdom, auditory and other sensory learning capabilities could considerably vary across species. However, since sensory learning capabilities can be associated with behaviors that are not tied to vocal production, an important question is: how to measure these abilities systematically and in ways that allow cross species comparisons?

Artificial-Grammar Learning (AGL) paradigms (Reber, 1967) are useful for understanding how different individuals learn the structure of a sequence of sensory elements. Artificial Grammars (AG) can be designed to create different levels of structural complexity in how elements are organized in a sequence. The learning of these sequences can be measured using non-vocal motor output (e.g., Fitch and Hauser, 2004; Gentner et al., 2006; Murphy et al., 2008). Generally, these experiments involve an initial phase where the animals are either explicitly trained to learn exemplary "correct" sequences that follow the AG structure, or they are habituated to the exemplary AG sequences. The latter approach aims to tap into more implicit forms of learning, similar to the way that infants glean the statistical properties of language-related structure (Saffran et al., 1996; Marcus et al., 1999). Subsequent to the learning phase, the animals are tested with novel "correct" and "violation" sequences to determine if they can distinguish them, either by their trained or natural responses (e.g., by measuring preferential looking responses towards the different testing sequences). As examples of the types of structures that can be studied with AGL paradigms, AGs can be designed to have only adjacent relationships between the elements in a sequence (Saffran et al., 1999; Fitch and Hauser, 2004; Friederici, 2004; Friederici et al., 2006), non-adjacent relationships between more distantly associated elements (Friederici et al., 2006; Pallier et al., 2011; Petersson et al., 2012), and/or hierarchically organized relationships (Bahlmann et al., 2008, 2009; Friederici, 2011). For further details on the historical basis for and the use of AGL paradigms in adult humans, infants or other animals see: (Reber, 1967; Fitch and Hauser, 2004; Fitch and Friederici, 2012; Petkov and Wilson, 2012).

In a few studies with songbirds, where starlings (Gentner et al., 2006) or Bengalese finches (Abe and Watanabe, 2011) participated in AGL paradigms, it was claimed that these species can learn hierarchically nested grammatical structures. However, these interpretations have been challenged on the grounds that it remains possible that the animals could have learned the difference between "correct" and "violation" sequences by using simpler strategies, which is considered in detail elsewhere (van Heijningen et al., 2009; Berwick et al., 2011; Ten Cate and Okanoya, 2012). Thus, some authors have concluded that it remains controversial whether any non-human animal can recognize auditory patterns that require grammars hierarchically higher than FSGs or regular grammars (e.g., context-free grammars, see Berwick et al., 2011).

Tamarins, a New World monkey species, seem able to perceptually learn adjacent relationships between FSG sequences (e.g., Fitch and Hauser, 2004), although it is not clear if this extends to the learning of non-adjacent relationships (also see: Newport et al., 2004). However, a number of the results on the testing of AGL in non-human primates that have used preferential looking paradigms to measure behavioral responses, have been questioned in part because of the subjective nature of experimenters rating the responses of animals captured on video (Ten Cate and Okanoya, 2012). Wilson and colleagues have devised some solutions to automate the analysis of natural eye-movement responses using non-invasive eye-tracker systems (Wilson et al., 2011). With this approach they have obtained evidence that Rhesus macaques can learn an auditory artificial-grammar with several forward branching relationships, such as those often seen in the produced songs of songbirds and cetaceans (Hurford, 2012). With greater objectivity, it is important to revisit the issue of what level of structural complexity in auditory pattern learning different animals are naturally capable (Petkov and Wilson, 2012).

Regarding what non-human primates are capable of learning with training, an interesting recent report trained baboons on pairwise associations between several visual symbols, e.g., A1-B1, A2-B2, etc. (Rey et al., 2012). In a later testing phase, the animals were presented with the initial "A" elements of two pairs (e.g., A1-A2) and were then allowed to select the "B" elements that would follow. Here, the animals were seen to preferentially pair the "B" partner of the most recent "A" element that was seen (e.g., A2-B2), followed by the partner pair of the first element (e.g., A1-B1). This resulted in the most often selected pattern, A1-A2-B2-B1, which resembles a hierarchical "center-embedded" (or nested) structure. It is interesting that the baboons seemed to rely on an associative memory trace of the pairs of elements that they were trained to recognize, which as the authors interpret may have had an evolutionary basis for human abilities to nest syntactic expressions. However, since FSG are subsets of hierarchically higher grammars and FSGs can generate sequences that can appear to be nested, whether the baboons can learn center-embedded structure remains unclear. Some linguists have outlined a set of criteria on which the animal work would need to be evaluated, if this is the objective (Jaeger and Rogers, 2012). Thereby, as with the related songbird studies (Gentner et al., 2006; Abe and Watanabe, 2011), it is currently unclear whether any non-human animal can learn patterns above those that can be generated by FSGs (or regular languages) in the FLH.

A need for continuing revision of the Formal Language Hierarchy combined with further comparative testing

Given that vocal learning and sensory learning capabilities appear to be more variable among vertebrates than is often appreciated (**Figure 1**), approaches in linguistics and those that rely on AGL paradigms remain useful for clarifying the extent of animal capabilities. However, there are important issues that tend to get overlooked which can limit our understanding of the structure of animal vocalizations or the extent of animal AGL capabilities:

- FSGs are subsets of languages higher on the FLH. Thus it is not always easy to know whether the vocal production or sensory learning of a particular set of sequences requires a higher-level process. Without evidence for a higher-level process a simpler process might be possible both in humans and other animals. For instance, humans can rely on semantics to simplify the complexity of a syntactic process and even humans can find AGL void of semantic content challenging to learn (Perruchet and Rey, 2005; Uddén et al., 2012).
- There has been considerable interest in understanding how high humans and other animals can reach into the FLH. However, by focusing solely on the top end of the FLH, the animal AGL experiments have tended to under-support some of the other potentially interesting aspects in the data on animal AGL. For example, it remains unclear the extent to which non-human animals can learn non-adjacent relationships between sounds, which many view as a key evolutionary transition in the evolution of human syntactic abilities (for a review: Fitch and Friederici, 2012).
- There are considerable levels of structural complexity in FSGs (Reber, 1967; Petersson et al., 2012) that need to be better resolved so that different types of AG structures can be systematically changed and/or compared to others (Hurford, 2012; Petkov and Wilson, 2012).

Some groups have been considering how the FLH can be resolved in greater detail (see: Hurford, 2012; Jaeger and Rogers, 2012; Petkov and Wilson, 2012). For example, Petkov and Wilson (2012) note that the simplest scenario for auditory learning is the recognition of a single sound/element, such as the recognition of a single vocalization from a limited set of vocalizations. With the recognition of two types of elements in a sequence, it is known that many animals habituate to the repetition of the same element and dishabituate to the introduction of a novel element (e.g., repetition effects, Grill-Spector et al., 2006). With three or more different elements, there is the possibility of creating a greater number of structural relationships in the transitions between elements. Continuing efforts are needed to quantify the multidimensional space of "syntactic complexity," especially for FSG structures where animal abilities vary. A better understanding of the graded levels of "syntactic complexity" in vocal production and sensory learning capabilities across species could clarify the origins of syntax and spoken language.

EVOLUTIONARY HYPOTHESES ON VOCAL AND AUDITORY LEARNING: GAINS, LOSSES OR EVERYONE HAS IT?

Phylogenetic comparisons suggested that complex-vocal learning evolved among birds at least two, if not three independent times: in oscine songbirds, parrots, and hummingbirds (Nottebohm, 1976; Jarvis et al., 2000; Hackett et al., 2008; Suh et al., 2011). The difference in the number of independent vocal learning events depends on the interpretation of different phylogenetic trees (**Figure 2**): (1) either three gains in all three lineages based on phylogenetic trees that are separated by multiple non-learners (Sibley and Ahlquist, 1990; Jarvis, 2004) or (2) two gains, in hummingbirds and the common ancestor of parrots and oscine songbirds, with a loss in the suboscine songbirds (Suh et al., 2011). To explain either of these observations, Jarvis (2004) proposed at least three not mutually exclusive hypotheses for the evolution of vocal learning: (1) complex vocal

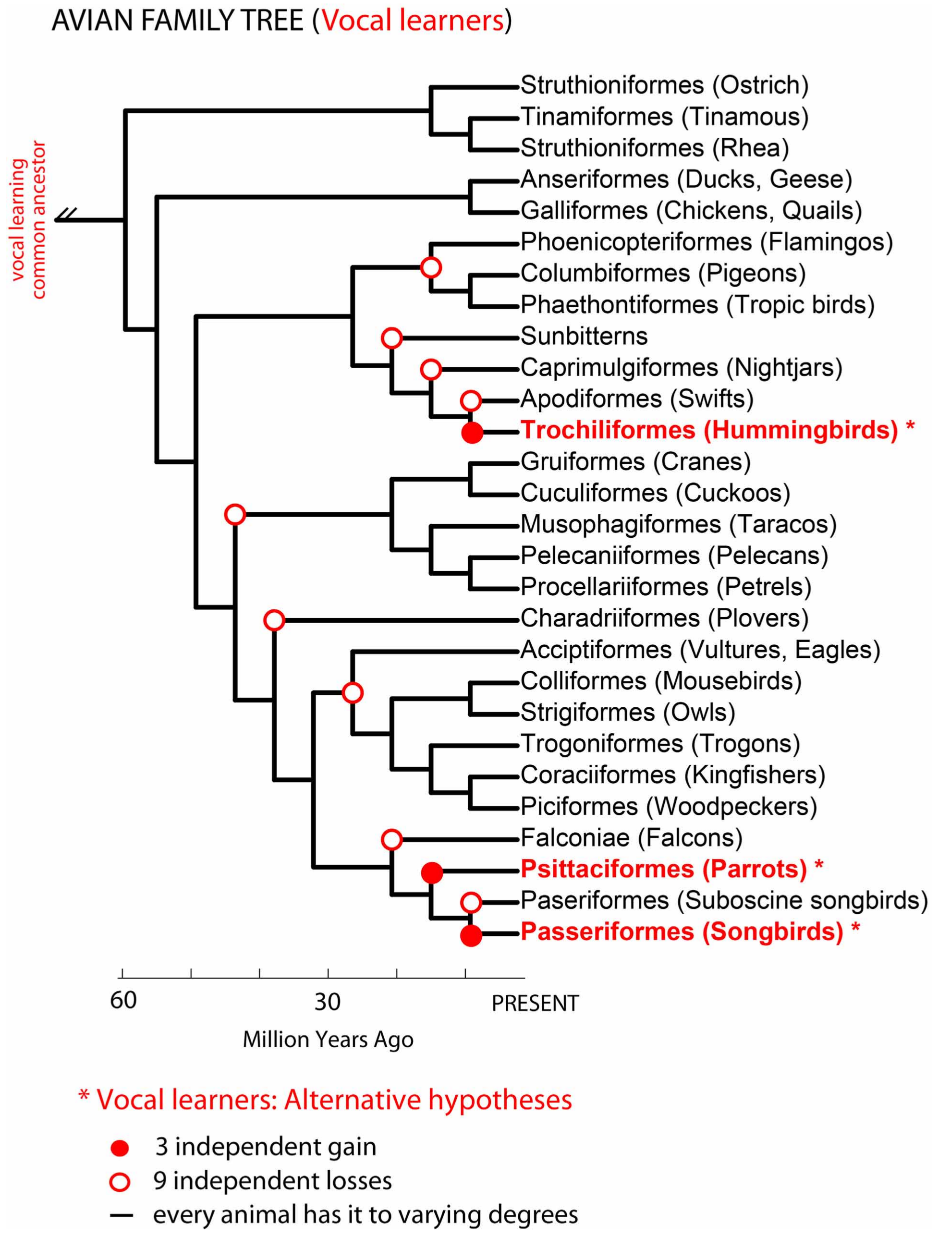


FIGURE 2 | Avian phylogenetic tree and the complex-vocal learning phenotype. Shown is an avian phylogenetic tree (based on: Hackett et al., 2008). Identified in red text and * are three groups of complex-vocal learning birds. Below the figure are summarized three alternative hypotheses on the evolutionary mechanisms of complex-vocal learning in birds (see text, and Jarvis, 2004). The auditory sequence learning phenotype described in **Figure 1B**, is not shown here, since some forms of auditory learning seem to be present in all birds. However, further comparative data is needed on the learning of the complexity of auditory sequences, which to our knowledge has been tested using Artificial Grammars only in songbirds (Gentner et al., 2006; van Heijningen et al., 2009; Abe and Watanabe, 2011).

learning independently evolved multiple times in birds; (2) complex vocal learning was lost either four (Jarvis, 2004) or nine times (Suh et al., 2011); and/or (3) all species are vocal learners to some extent. We note that vocal learning being independently gained or lost suggests a categorical distinction between vocal learners and vocal non-learners. Vocal learning being more continuously distributed among many species than categorical would indicate that gains and losses can occur to a greater extent.

In the primate phylogenetic tree, only humans are thought to be complex-vocal learners (**Figure 3** solid red circle). As with birds, one possible evolutionary hypothesis is that humans evolved vocal learning independently from other primates. Alternatively, if we suppose that a primate ancestor was a complex vocal learner, complex-vocal learning would have to have been lost at least eight times in the primate order (**Figure 3**, open red circles) and maintained in humans. The evolutionary losses hypotheses become less tenable when the number of losses greatly exceeds the number of independent gains. Putting this together, according to these phylogenies and vocal phenotypes, the number of independent gains is: 1 in primates (**Figure 3**), 2–3 in birds (Feenders et al., 2008), and 5 in mammals including humans (Jarvis, 2004). The number of losses can be as high as: 8 in primates (**Figure 3**), 4 or 9 in birds (Feenders et al., 2008; Hackett et al., 2008), and 11 in mammals (Fitch and Jarvis, in press). If the losses are true, what could explain such high rates of losses? One

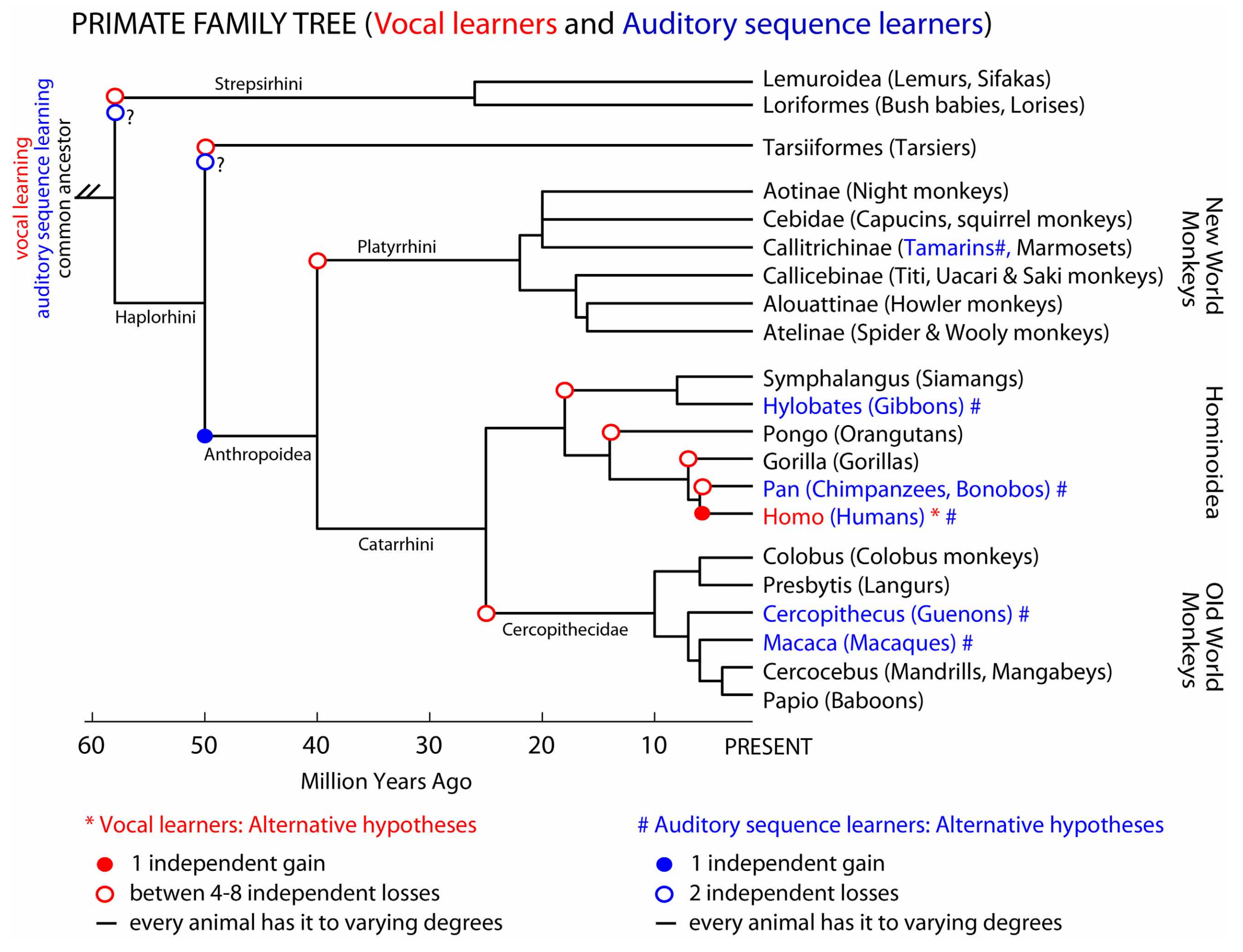


FIGURE 3 | Primate phylogenetic tree and complex-vocal learning vs. auditory sequence learning. Shown is a primate phylogenetic tree based on a combination of DNA sequence and fossil age data (Goodman et al., 1998; Page et al., 1999); for a recent review see (Cartmill, 2010). Humans (Homo) are the only primates classified as "vocal learners." However, non-human primates might be better at auditory sequence learning than their limited vocal-production learning capabilities would suggest. In blue text and (#) we highlight species for which there is some evidence of Artificial Grammar Learning capabilities for at least adjacent relationships between the elements in a sequence (tamarins: Fitch and Hauser, 2004), (macaques: Wilson et al., 2011). Presuming that the auditory capabilities of guenons and gibbons mentioned in the text (or the symbolic learning of signs by apes) would mean that these animals are able to learn at least adjacent relationships in Artificial Grammars we can tentatively mark these species also in blue #. Note however, that for the species labeled in black text, future studies might show them to be capable of some limited-vocal learning or various levels of complexity in learning the structure of auditory sequences. Three not mutually exclusive hypotheses are illustrated for both complex-vocal learning and auditory sequence learning.

idea is that predatory influences may have selected against vocal learning by selecting against complex vocalization sequences that would allow predators to better localize their prey (Hosino and Okanoya, 2000; Jarvis, 2004, 2006). Some support for this notion is that the known mammalian vocal learners (humans, elephants, and cetaceans) are at or near the top of the food chain, and some of the avian vocal learners (corvid songbirds, hummingbirds, and parrots) are considered exceptional at escaping predators (Jarvis, 2006). Nonetheless, the evolutionary mechanisms may not necessarily be the same across animal species.

Similar forms of gains, losses or other hypotheses could be applied to auditory (sequence) learning abilities. However, here there is a greater paucity of comparative data. Non-human animals may considerably differ in their ability to learn the various levels of sequencing complexity in AG, which at face value could be considered to have evolved independently or by common descent (**Figure 3**). In several non-human primate species there is an impression of at least the ability to learn adjacent auditory relationships in AG structures (**Figure 3**, blue nodes and text). Some of these species have also been shown to have relatively simple combinatorial production capabilities (Arnold and Zuberbuhler, 2006). However, we are not aware of evidence for or against prosimians (lemurs, bush-babies, etc.) being able to perceptually learn various levels of structural complexity in AGs or to produce simple sequences with their vocalizations. Thus, additional comparative study is needed to fill in this currently tentative picture (**Figure 3**). In this regard, as we have argued, developments in linguistic theory and AGL approaches can help us to characterize the extent of the syntactic-like capabilities of non-human animals either for production or sensory learning. We further argue that understanding the distinctions in such behavioral phenotypes and their mechanisms across species will require an improved understanding of their neurobiological substrates.

NEUROBIOLOGICAL PATHWAYS FOR VOCAL PRODUCTION

Humans heavily rely on a forebrain pathway to produce learned vocalizations. This pathway is thought to be in many ways separate from an ancestral pathway in non-human primates for producing innate vocalizations (Jurgens, 2002; Jarvis, 2004). Similarly, complex-vocal learners such as songbirds, parrots, and hummingbirds have distinct vocal learning forebrain nuclei that have so far not been found in other birds. That is, for birds, despite the noted variability in the behavioral evidence for vocal learning (**Figure 1**), the published neurobiological evidence has highlighted distinctions between the neurobiological substrates for vocal production in so-called vocal learners and vocal non-learners (**Figures 4**, **5A,B**). We overview this literature here, which might be challenged or supported by future work.

DIFFERENT SUBSYSTEMS AND DIRECT AND INDIRECT PATHWAYS FOR VOCALIZATION IN PRIMATES

Historically, our understanding of the brain pathways involved in the production of innate vocalizations in primates stems from the classical brain stimulation studies of Penfield and colleagues (Penfield and Rasmussen, 1949) and the anatomical studies of Kuypers in human and non-human primates (Kuypers, 1958a,b, 1982). These studies were followed by over 40 years of work by several groups in different monkeys (such as squirrel monkeys and macaques) using anatomical tracing, neurophysiological recordings during vocalization, lesions to affect vocalizations, and microstimulation to either elicit vocalization or to contract laryngeal muscles (for reviews: Jurgens, 2002, 2009; Simonyan and Horwitz, 2011). Others compared the results in primates (Hast et al., 1974) with those from cats and dogs (Milojevic and Hast, 1964), suggesting some key differences between the human, monkey, and carnivore vocal production subsystems.

Primates are thought to have a limbic (affective) or innate vocal-production subsystem (**Figures 4B,D**) that involves connections from the amygdala, orbito-frontal cortex and anterior-cingulate cortex to the periaqueductal gray (PAG) in the brainstem (Jurgens and Pratt, 1979b,a; Kirzinger and Jurgens, 1982, 1985). Neurons in the PAG synapse onto neurons in the reticular formation, which in turn synapse with the α-motoneurons in the nucleus ambiguus (Dujardin and Jurgens, 2005). The motoneurons in the nucleus ambiguus control the muscles of the larynx for vocal production (**Figures 4B,D**). The PAG and reticular formation are required for the motor production of vocalizations via nucleus ambiguus (Jurgens, 2002; Hage and Jurgens, 2006; Hannig and Jurgens, 2006).

Non-human primates also have a ventro-rostral cortical region in Brodmann Area 6 (Area 6vr) that projects to the vocal production areas of the reticular formation, which in turn projects to nucleus ambiguus (Simonyan and Jurgens, 2003), see **Figure 4D**. When this area in non-human primates is stimulated, it contracts the laryngeal muscles (Hast et al., 1974). Area 6vr and the anterior cingulate are also interconnected with parts of the primary motor cortex, amygdala, and ventral and midline thalamus. However, when Area 6vr is stimulated no vocalizations are produced (Hast et al., 1974; Simonyan and Jurgens, 2003), and when it is lesioned vocalizations are reportedly unaffected (Jurgens, 2002). Thus, it has been proposed that Area 6vr controls respiration associated with vocalization rather than the control of vocalization (Jurgens, 2002, 2009).

However, there is growing evidence that, in non-human primates, neurons in Area 6vr or surrounding regions in the ventral prefrontal/premotor cortex of non-human primates can be modulated during innate vocalization production; although at this point it is not clear whether the results depend on the animals hearing their own vocalizations or the context surrounding the vocalizations. A recent study in macaques suggests that when monkeys produce innate vocalizations on cue, some neurons show premotor neural activity in a frontal cortical area near or in Area 6vr (Coude et al., 2011). Notably, the animals made the vocalizations during a learned non-vocal task and the activity response only occurred under certain conditions. Two other studies in common marmoset monkeys (*Callithrix jacchus*) reported on neuronal activity-related gene expression patterns during vocalization. In one of these studies higher numbers of *egr*-1 immunopositive cells were observed in the prefrontal cortex when the animals vocalized relative to when they remained silent (Simões et al., 2010). In the other study, *c-fos* induction was higher in the ventrolateral prefrontal cortex during vocal perception and production ("anti-phonal" calling) than during vocal production

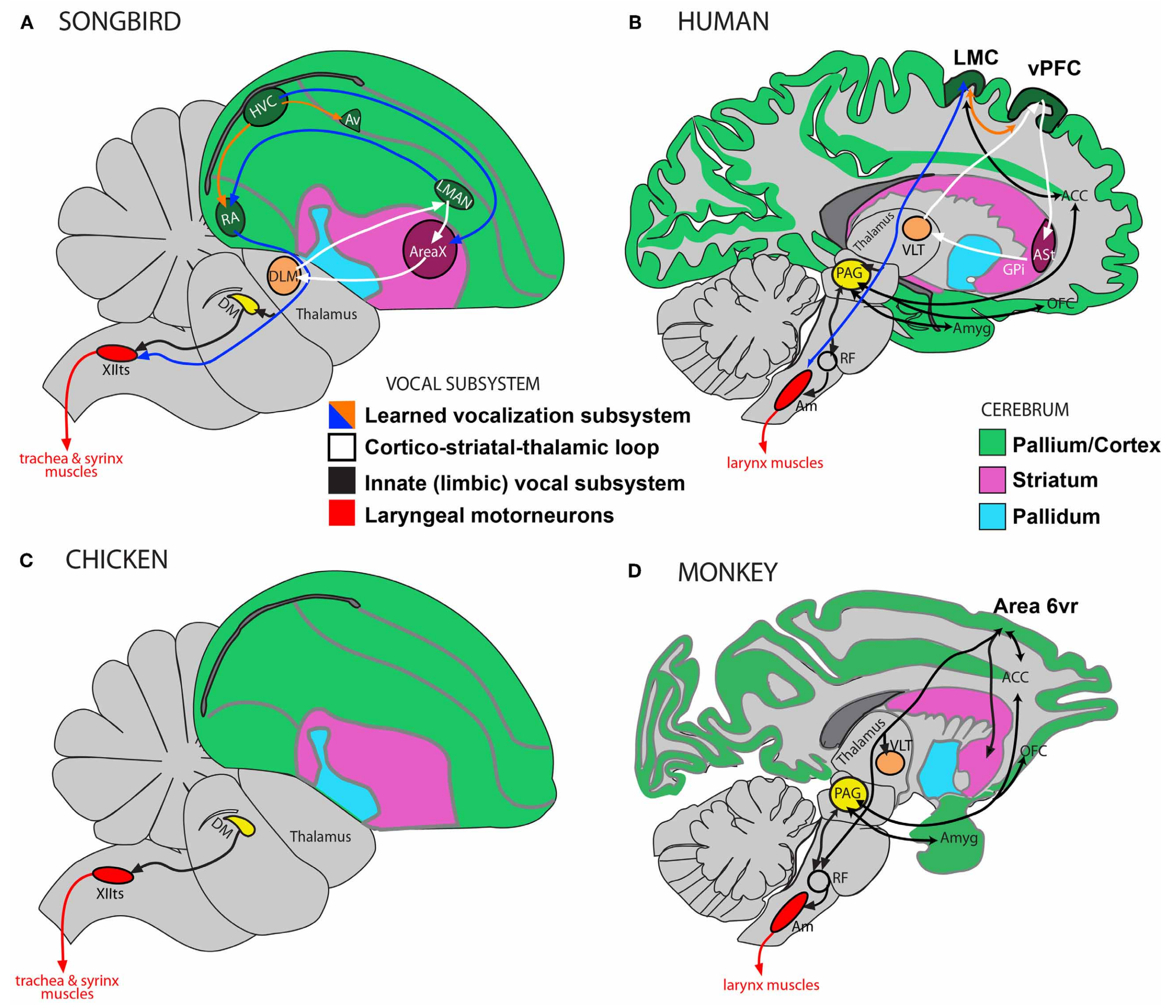


FIGURE 4 | Vocalization subsystems in complex-vocal learners and in limited-vocal learners or vocal non-learners: Direct and indirect pathways. The different subsystems for vocalization and their interconnectivity are illustrated using different colors. **(A)** Schematic of a songbird brain showing some connectivity of the four major song nuclei (HVC, RA, AreaX, and LMAN). **(B)** Human brain schematic showing the different proposed vocal subsystems. The learned vocalization subsystem consists of a primary motor cortex pathway (blue arrow) and a cortico-striatal-thalamic loop for learning vocalizations (white). Also shown is the limbic vocal subsystem that is broadly conserved in primates for producing innate vocalizations (black), and the motoneurons that control laryngeal muscles (red). **(C)** Known connectivity of a brainstem vocal system (not all connections shown) showing absence of forebrain song nuclei in vocal non-learning birds. **(D)** Known connectivity of limited-vocal learning monkeys (based on data in squirrel monkeys and macaques) showing presence of forebrain regions for innate vocalization (ACC, OFC, and amygdala) and also of a ventral premotor area (Area 6vr) of currently poorly understood function that is indirectly connected to nucleus ambiguous (see text). The LMC in humans is directly connected with motoneurons in the nucleus ambiguus, which orchestrate the production of learned vocalizations (also see **Figure 5B**). Only the direct pathway through the mammalian basal ganglia (ASt, anterior striatum; GPi, globus palidus, internal) is shown as this is the one most similar to AreaX connectivity in songbirds. Modified figure based on (Jarvis, 2004; Jarvis et al., 2005). Abbreviations: ACC, anterior cingulate cortex; Am, nucleus ambiguus; Amyg, amygdala; AT, anterior thalamus; Av, nucleus avalanche; DLM, dorsolateral nucleus of the medial thalamus; DM, dorsal medial nucleus of the midbrain; HVC, high vocal center; LMAN, lateral magnocellular nucleus of the anterior nidopallium; LMC, laryngeal motor cortex; OFC, orbito-frontal cortex; PAG, periaqueductal gray; RA, robust nucleus of the of arcopallium; RF, reticular formation; vPFC, ventral prefrontal cortex; VLT, ventro-lateral division of thalamus; XIIts, bird twelfth nerve nucleus.

alone, which had higher *c-fos* induction in dorsal premotor cortex (Miller et al., 2010). These findings suggest that the sensory input into these regions may be an important factor for neuronal activation. It has thus become important to determine whether sensory input or motor activity during vocalization is primarily responsible for the observed results by temporarily deafening the vocalizing animals with ear plugs or some other ethically acceptable manipulation. Such future work is likely to clarify the functional role of the prefrontal/premotor cortex, including Area 6vr, during vocalization in non-human primates.

Humans are thought to also rely on the innate vocal-production pathway (e.g., cingulate, amygdala, PAG, and nucleus ambiguus) for producing involuntary vocalizations, such as, when a person shrieks to an aversive stimulus. However, humans rely considerably more on another pathway for learned vocalization, i.e., speech production. This pathway includes the primary motor cortex, regions in the lateral inferior and middle frontal cortex, premotor cortex, supplementary motor cortex, cerebellum, and subcortical structures such as, parts of the basal ganglia and thalamus (Jurgens, 2002; Jarvis, 2004; Simonyan and Horwitz, 2011). In humans, this so-called learned vocal pathway appears to have a direct projection from the face area of primary motor cortex in Brodmann Area 4 to the nucleus ambiguus (Kuypers, 1958a; Iwatsubo et al., 1990; Jurgens, 2002; Khedr and Aref, 2002). This human region in BA4 is also called the Laryngeal Motor Cortex (LMC; **Figure 4B**). When it is stimulated the participants vocalize (reviewed in: Simonyan and Horwitz, 2011). When human LMC has been damaged the production of learned vocalizations is eliminated. No such homolog of the LMC region has been found in the primary motor cortex of non-human primates, either with stimulation or by lesion. Homologs of human LMC or the non-human primate Area 6vr have been searched for in some non-primate mammalian species, such as cats and dogs, but have not been found (Milojevic and Hast, 1964).

Based on these findings, many investigators hypothesized that the evolution of spoken language in humans was associated with the formation of a direct projection from LMC to nucleus ambiguus in humans (Kuypers, 1958a; Kirzinger and Jurgens, 1982; Iwatsubo et al., 1990; Jurgens, 2002; Khedr and Aref, 2002; Jarvis, 2004; Okanoya, 2004; Fitch et al., 2010; Fischer and Hammerschmidt, 2011; Simonyan and Horwitz, 2011). Simonyan and Horwitz (2011) (also see: Simonyan and Jurgens, 2003), hypothesized that the *function* of the Area 6vr region migrated from its presumed ancestral premotor cortex location in non-human primates into the primary motor cortex to become the LMC in humans, simultaneously taking over direct control of the nucleus ambiguus. To test this hypothesis, further work is needed to clarify whether humans have an area with the functionality and connectivity of non-human primate Area 6vr for contracting the laryngeal muscles. Also, the functional significance of the direct projection in humans from LMC to the nucleus ambiguus remains unclear, relative to the indirect projection from Area 6vr in other primates. Direct motor cortex control of motorneurons controlling hand and finger movement is seen to various extents in both human and non-human primates but less so in rodents (Lemon, 2008). However, a recent finding in laboratory mice appears to have revealed an LMC-like region (Arriaga et al., in press), which is active by vocalization production and makes a direct, but very sparse, projection to nucleus ambiguus, also see: (Arriaga and Jarvis, in press). This finding motivates a re-evaluation of the origins of the LMC in humans. In this regard, although a number of studies state that in non-human primates there is an absence of a direct projection from motor cortex to nucleus ambiguus, Kuypers' original 1958b study mentioned finding some peri-central cortical axons in the nucleus ambiguus of monkeys and chimpanzees. These findings can support the continuum hypothesis of vocal learning: Arriaga and Jarvis (in press) hypothesize that in addition to the presence vs. absence of the direct nucleus ambiguus projection, the density of the projection would be correlated with the level of limited to more complex vocal learning.

DISTINCT VOCAL FOREBRAIN NUCLEI IN COMPLEX-VOCAL LEARNING BIRDS

The differences in neural pathway connectivity for complex-vocal learning and limited-vocal learning or vocal non-learning birds are seen to parallel some of the findings in mammals. Best studied in songbirds and parrots, the neurobiological substrates for vocal control and learning includes four nuclei in an anterior forebrain pathway loop (which interconnect the pallium with the basal ganglia and thalamus) and three in a posterior pathway of which the robust nucleus of the arcopallium (RA) makes a direct projection onto the vocal motoneurons in the twelfth nerve nucleus (XIIts) of the bird brainstem, which control the muscles of the syrinx (**Figure 4A**, also: Jarvis, 2004; Bolhuis et al., 2010). No such nuclei or direct projections have been found in so-called vocal non-learning birds, such as ducks and pigeons (Wild, 1997; Dubbeldam, 1998; Jarvis, 2004). This direct projection is reminiscent of the direct projection in humans from LMC to the nucleus ambiguus that appears to be absent in non-human primates. All birds studied to date, however, have been shown to have brainstem input from the midbrain region DM (dorsal medial nucleus of the midbrain) to XIIts (**Figures 4A,C**), which, like the mammalian PAG projection to the nucleus ambiguus (via the reticular formation), controls the production of innate vocalizations. These cross species differences have been used to strengthen the hypothesis on the evolution of the direct projection being crucial for the evolution of vocal learning (Wild, 1997; Jarvis, 2004; Fitch et al., 2010).

To gain insights into the evolutionary bases of the vocal learning nuclei, Feenders and colleagues (2008) compared the forebrain vocal nuclei and adjacent brain regions in animals from each complex-vocal learner lineage—songbirds, parrots, and hummingbirds—to so-called vocal non-learners such as doves or non-singing female songbirds.[1] Extending prior studies (Jarvis and Nottebohm, 1997; Jarvis and Mello, 2000; Jarvis et al., 2000) the authors made the following key observations: First, when vocal learning birds performed non-vocal movement behaviors, such as hopping and flying, expression of the *egr*1 immediate early gene (associated with increases in neuronal activity) was restricted to forebrain regions surrounding or directly adjacent to the forebrain song learning nuclei. Second, in the vocal non-learning birds, comparable activated regions in non-vocal movement areas were found, but without the presence of forebrain song nuclei adjacent to them. Third, the activity-dependent gene activation in these regions was motor-driven and

---

[1] It is interesting that in some songbirds, only the males learn to sing. In these species, the females are born with song nuclei, but these atrophy as the females become adults, constituting a developmental loss of the trait. Delivery of high levels of estrogen in nestlings can prevent the atrophy of the song learning nuclei in these females (Konishi and Akutagawa, 1988). Hormonal manipulations in "vocal non-learners" do not seem to reveal a suppressed vocal system.

was independent of at least auditory or visual input. The *egr*1 expression in the movement-activated regions was correlated with the amount of body movements (e.g., wing beats) performed, whereas in the song-learning nuclei it was correlated with the amount of singing performed. Lastly, both the vocal learners and non-learners were found to have forebrain auditory pathways that are activated when the animals heard vocalizations, and with no noted differences between vocal learners and non-learners.

Feenders and colleagues (2008) used these results to propose a motor-theory of vocal learning origin. They propose that the brain regions in the vocal learning pathway derived from the same cell lineages that gave rise to the motor pathway for movement control unrelated to vocal production in birds. They propose that the new pathway then formed a direct projection onto the brainstem vocal motor neurons for greater control of vocal production. However, the equivalent function of the non-human primate Area 6vr (which when stimulated contracts laryngeal muscles) and its indirect projection to the vocal motoneurons, has yet to be found in so-called vocal non-learning birds (compare **Figure 4C** in chickens to **4D** in monkeys).

Other factors have been proposed to differ between vocal learners and non-learners, a common factor being hemispheric lateralization. It is known that in both humans and song learning birds there is a dominant hemisphere for learning, production, and processing of vocalizations, being left dominant in humans and canaries, and right dominant in zebra finches (Nottebohm et al., 1976; Simpson and Vicario, 1990; Phan and Vicario, 2010). Some have suggested that the stronger engagement of the left hemisphere in human language processing was a recent evolutionary adaptation (Tyler et al., 2011). This predicts a more bilateral engagement in the brains of limited-vocal learners that are closely evolutionarily related to humans. However, although lateralized functions for non-vocal behaviors have been seen in many species (Halpern et al., 2005), lateralized processing of communication signals in non-human primates, for instance, is sometimes (Heffner and Heffner, 1984; Poremba et al., 2004; Joly et al., 2012) but not always seen or explicitly tested for (for a review see: Petkov et al., 2009). Although lateralization is not restricted to humans, or to vocal learners, the question that remains is whether the level of lateralization, rather than the particular hemisphere, might be the critical variable for differences between complex-vocal learning and other species (Teufel et al., 2010). The ability to simultaneously image both hemispheres in birds, primates and other animals (e.g., Petkov et al., 2006; Boumans et al., 2008; Poirier et al., 2009; Baumann et al., 2011) can provide data for testing hemispheric effects.

SUMMARY OF VOCAL PRODUCTION PATHWAYS IN BIRDS AND PRIMATES

We saw in sections "Different Subsystems and Direct and Indirect Pathways for Vocalization in Primates" and "Distinct Vocal Forebrain Nuclei in Complex-Vocal Learning Birds" that primates and birds appear to share a broadly conserved pathway for producing innate, emotionally or spontaneously driven vocalizations. However, humans and song-learning birds appear to rely considerably more on a forebrain motor system for learned vocalization. The learned vocal-production subsystem has different connectivity with the motor neurons of the laryngeal (in mammals) or syringeal (in birds) muscles than the innate vocal-production subsystem. In all birds, the adjacent forebrain pathway appears to orchestrate motor action unrelated to vocal production, such as, wing flapping or hopping, both of which require movement coordination (Feenders et al., 2008). The same might be the case for primates although this is currently unknown. Various authors (Farries, 2004; Jarvis, 2004; Feenders et al., 2008) have suggested that the simplest evolutionary mechanism for vocal learning is that a genetic mutation established the link between the newly evolved forebrain nuclei and the vocal motor brainstem nucleus for vocal production (compare **Figure 4A** in songbirds to **4C** in chickens). In sections "Summary of Motor and Other Theories" and "Predictions of Motor and Other Theories, From a Modified Behavioral Perspective" we consider this and other, not mutually exclusive, hypotheses, which is based on re-evaluation of motor and cognitive theories that make different predictions about the neurobiological systems for production and perceptual learning.

AUDITORY INPUT INTO THE VOCAL PRODUCTION PATHWAYS

Because auditory learning is necessary but not sufficient for vocal learning, one might expect the auditory pathways to provide input into the vocal learning system in the complex-vocal learners but perhaps not for animals that are obligate vocal non-learners. Such auditory input has been the topic of extensive investigation in songbirds and parrots, but without yet a clear resolution (Jarvis, 2004; Mooney, 2009; Margoliash and Schmidt, 2010). In songbirds, the forebrain auditory pathway provides input into the interfacial nucleus of the nidopallium (NIf; a song nucleus) including the high vocal center (HVC) shelf and RA cup regions adjacent to the vocal motor pathway nuclei HVC and RA (**Figures 4**, **5**). The shelf and cup in turn are thought to send weak projections into HVC and RA, whereas NIf sends a strong projection into HVC (Vates et al., 1996; Jarvis, 2004; Mooney, 2009; Yip et al., 2012). Relatedly, human neuroimaging studies have described auditory cortex input into the frontal speech production areas (e.g., Rauschecker and Scott, 2009; Friederici, 2011; Tyler et al., 2011).

In so called vocal non-learners or limited vocal learners, there is considerable data on the structure and function of the auditory pathway from cochlea to cortex, including in non-human primates and other vertebrates (e.g., Rauschecker, 1998; Carr and Code, 2000; Kaas and Hackett, 2000) and song learning birds and pigeons (Mooney, 2009; Margoliash and Schmidt, 2010). In all of these sets of species, the auditory pathway projects from the cochlea to the midbrain auditory nucleus, to the thalamic auditory nuclei, and then to primary and secondary auditory cortical/pallial regions. After entering the forebrain, in vocal non-learners auditory input is thought to enter motor pathways, but in the complex-vocal learners it also enters the vocal motor pathways. If the presumed vocal non-learners are thought to primarily rely on an innate vocal-production system, then auditory input into the vocal production system would not seem to be required for genetically regulated vocal production.

To clarify the neurobiological substrates for auditory processes, vocal production learning, and the interface of the two, it has become critical to: (1) determine which animals are strictly vocal non-learners; (2) whether the neurobiological vocal production pathways in complex-vocal learners are as clearly distinct from those of limited-vocal learners as they seem; and (3) if there are differences across the species in the dependence of the vocal production subsystems on input from the auditory system. These clarifications are needed because the distinctions between "vocal learners" and "vocal non-learners" in their sensory-motor (e.g., auditory-vocalization) interactions are at the core of certain motor, gestural and cognitive theories.

SUMMARY OF MOTOR AND OTHER THEORIES

Motor theories are appealing for explaining sensory-motor relationships in communication for the following reasons. The sequencing of motor behaviors at multiple scales is an ancestral function. For example, many quadruped mammals increase their speed of movement by shifting from a walking gait to a running gait, each requiring different coordinated sub-movements of the limbs and sensory-motor feedback (Schmitt, 2010). Human language involves the temporal sequencing of laryngeal and other oral-facial muscles, and respiratory apparati, to produce speech sounds at multiple temporal sequencing levels, including phonological, sub-lexical and lexical, and syntactic. These forms of sequencing are used for perception and production. In the case of language syntax perception, humans often evaluate hierarchically organized dependencies between words in a sentence that cannot be simply solved by sequentially evaluating the words (Bickerton, 2009). Language production also requires coordinating a series of muscle movements of the larynx with feedback from the sensory system. Thereby spoken-language perception and production depend on sensory-motor interactions and these are differently emphasized by the various theories.

Although there are several motor theories in the literature, in this section we compare two sets of not mutually exclusive theories: motor theories of speech/song perception (Liberman and Mattingly, 1985; Williams and Nottebohm, 1985), and a motor theory of vocal learning origin (Feenders et al., 2008). As variants of motor theories, we briefly overview the "gestural theory of spoken language evolution" (Hewes, 1973) and the "gestural (mirror neuron) hypothesis of language evolution" (Rizzolatti and Arbib, 1998). Then we compare them with alternatives to motor/gestural theories, namely broadly conserved "sensorimotor integration" and "cognitive domain general" hypotheses.

MOTOR THEORIES OF SPEECH/SONG PERCEPTION

The well-known motor theories of speech perception in humans (Liberman and Mattingly, 1985) and song perception in songbirds (Williams and Nottebohm, 1985), make the strong claims that speech and song perception are primarily driven by the motor system. Although, one might expect the perception of speech sounds to be a perceptual problem for the auditory system, Lieberman and Mattingly argue that it is difficult to explain a large set of speech perception phenomena by only their sensory representation, since speech perception more often departs from its sensory representation than does the perception of other sounds. The theory proposes that the sensory-motor transformations made during speech perception and production are overlearned in humans. Because of this, the motor system actually drives auditory representation of speech to expedite the perception of speech in a way that is not available for the perception of other sounds.

Others have aimed to generalize the motor theory for speech perception to syntax perception. Allott suggested that the motor system would be important for the perceptual sequencing of syntactic expressions and for preparing syntactically organized sentences for production (Allott, 1992). An interesting variant of the motor theory of speech perception argues that the motor cortex is not necessary for speech perception, *per se*, but is necessary to sequence a conversation between two speakers, such as controlling when the speakers take turns in a conversation (Scott et al., 2009). The motor theory of song perception in songbirds as originally proposed was based on observations that the entire song learning system (from HVC to the descending pathway involving the vocal motoneurons in nucleus XIIts) shows song selective auditory responses (Williams and Nottebohm, 1985); for reviews see Mooney, 2009; Margoliash and Schmidt, 2010.

MOTOR THEORY OF VOCAL LEARNING ORIGIN ACROSS SPECIES

Similar to the motor theory proposed for vocal learning origin in birds (section "Distinct Vocal Forebrain Nuclei in Complex-Vocal Learning Birds"), the same authors proposed a similar theory for humans based on consideration of the evidence in the human literature (Feenders et al., 2008). Like in birds, the theory proposes that humans rely on a speech/song-learning pathway that is based on elaboration of a pre-existing motor pathway that controls learned movement sequencing. This would mean that vocal non-learning birds and mammals only have the forebrain motor pathway that supports movement patterning abilities unrelated to those for vocal production. By comparison, vocal learners evolved a new pathway in parallel to control the vocal motor neurons. In essence, in this theory, like mechanisms of gene evolution, the vocal learning pathway in birds and humans is seen as forebrain motor pathway duplication that adapted to directly control the muscles of the larynx/syrinx in addition to other muscle groups for respiration.

GESTURAL THEORIES

There are at least two independently developed gestural theories of language evolution: (1) The general "gestural theory of spoken-language origin" (Hewes, 1973; Tomasello et al., 1993); and (2) The gestural mirror neuron hypothesis of language evolution (Arbib, 2005; Prather et al., 2008; Arbib, 2010). The general gestural theory proposes that the brain pathways controlling the production of speech emerged from ancestral brain pathways controlling learned gestures. Thereby human and some non-human primates can perform learned gestures, but only humans can learn vocalizations relying on the gestural motor system. This theory is similar to the motor theory of vocal learning origin (Feenders et al., 2008). However, the two theories differ in that the gestural theory implies that the brain regions supporting gesturing and speech perception overlap, whereas, the motor theory implies that the more general movement control system was adapted for spoken language.

The gestural mirror neuron hypothesis tries to explain the evolutionary mechanisms of speech production learning by relying on "mirror neuron" results in primate and, more recently, avian vocal motor imitation (Arbib, 2005; Prather et al., 2008; Arbib, 2010). This theory was developed from the discovery in non-human primates that the same neurons fire both when the same action is observed or produced (di Pellegrino et al., 1992). Such neurons have been observed in frontal and parietal cortex. In humans, brain imaging has been used to localize regions presumably containing mirror neurons (e.g., Chong et al., 2008). The gestural mirror neuron hypothesis of language evolution argues that brain pathways that generate speech also process speech and visual gestures, and by doing so are able to transfer and copy hearing or seeing into motor behavior in the motor pathways. Non-human primates are said to have this system, but only for visual to motor copying of neural signals. Humans are said to have it for both visual and auditory to motor copying of signals, including for spoken language (Rizzolatti and Arbib, 1998; Arbib, 2005). The lack of the auditory-vocal motor link in non-human primates is thought to result from a lack of, or weakness of, a link in the auditory to vocal motor pathway, rather than the absence of a vocal motor pathway. In songbirds, Prather and colleagues have found a direct vocal motor link in mirror-neurons. They discovered neurons in the song nucleus HVC that have comparable responses to the production of learned songs and to hearing the songs (Prather et al., 2008). However, the relationship between auditory to vocal mirror neurons in the vocal motor pathway in primates remains largely theoretical (Rizzolatti and Arbib, 1998; Arbib, 2005, 2010).

SENSORIMOTOR INTEGRATION AND COGNITIVE "DOMAIN-GENERAL" HYPOTHESES

The above motor and gestural theories differ in the mechanisms of sensory-motor interactions, but share the notion that the auditory-motor interactions in humans and other vocal learning animals have specialized considerably relative to those in other animals. Rauschecker and Scott have proposed a "sensorimotor integration" model (Rauschecker and Scott, 2009; Rauschecker, 2011) that highlights the broadly conserved aspects of auditory-motor processing in human and non-human primates. This model builds on the notion of evolutionarily conserved auditory pathways in human and non-human primates (Romanski et al., 1999; Tian et al., 2001; Rauschecker and Scott, 2009) and other mammals (e.g., cats, Lomber and Malhotra, 2008). They propose that a ventral auditory pathway from the temporal lobe to ventral prefrontal cortex is engaged in processing auditory "objects" (such as calls in animals and speech in humans) and a dorsal auditory-to-premotor pathway for auditory-to-motor interactions that includes language-related processing in humans (for reviews see: DeWitt and Rauschecker, 2012; Rauschecker, 2012). The "sensorimotor integration" model proposes that the dorsal pathway receives efference copies from prefrontal and premotor regions that can affect auditory processing and perception. Such models emphasize the commonalities across the species regarding sensorimotor integration (Rauschecker and Scott, 2009; Rauschecker, 2011). An analogous, but also different efference copy model has been proposed for songbirds. Instead of an efference copy being sent back to the auditory pathway, the copy is thought to be sent from the song nucleus HVC to the anterior forebrain pathway through the basal ganglia for comparing planned motor output with auditory feedback (Troyer and Doupe, 2000; Fee, 2012).

There are also cognitive hypotheses, such those based on the notion that language processing involves "domain general" cognitive processes that have improved over evolution and also improve during child development (e.g., Saffran et al., 1996; Marcus et al., 1999; Friederici et al., 2011; Perani et al., 2011). This is in contrast to the notion that language involves domain specific modules that have specialized specifically for language processing. In support of domain general hypotheses, there is evidence that the processing of AG structures can engage comparable brain regions as does the processing of natural language material (Friederici et al., 2006; Bahlmann et al., 2008; Folia et al., 2011; Tyler et al., 2011; Petersson et al., 2012). Such hypotheses tend to emphasize the role of the cognitive systems, such as those supporting attention and memory, and how these may have improved during evolution to support language in humans or vocal learning in complex-vocal learners.

PREDICTIONS OF MOTOR AND OTHER THEORIES, FROM A MODIFIED BEHAVIORAL PERSPECTIVE

In this final section, we aim to integrate the ideas generated in the previous sections. We first summarize ways in which the prediction from the motor theories could be tested. Second, we summarize predictions from the other theories that we have considered, including cognitive domain-general hypotheses and how the predictions of these theories relate to those from the motor theories. The integration of the experimental and theoretical strands is important for advancing our understanding of language origins and mechanisms, and more generally of animal communication.

PREDICTIONS FROM MOTOR THEORIES

Strong theories make predictions that can be tested to help to support or refute their different tenets. We suggest that all such theories of spoken language evolution should be tested at both the behavioral and neural levels in order to revise them or to develop better ones. Next we consider the testing of predictions from the different theories in the context of our modified views, based on the accumulating evidence on variability in vocal production and auditory learning abilities in different species.

Predictions of motor theories of speech/song perception

These theories suggest that there are considerable benefits for the *perception* of conspecific communication signals in animals that can rely on learned sensory-motor interactions during vocal imitation or vocal learning. Thus, they predict considerable behavioral differences in the perception of learned sounds between so-called vocal learners and non-learners (or even limited-vocal learners). They also predict specialization in at least the sensory-vocal motor interconnectivity in complex-vocal learners that would be lacking or limited in more limited vocal learners. Because limited-vocal learning animals do not readily mimic others' vocalizations and apparently do not have a functional vocal motor forebrain pathway (Feenders et al., 2008), their sensory systems would not benefit from interaction with a forebrain vocal system for the perception of vocal communication signals.

Testing this theory depends on whether results are expected to show absolute differences (presence vs. absence) or differences by degree. Absolute differences between complex- and limited-vocal learners in the behavioral perception of communication signals and the neural substrates that subserve them are unlikely to be found in biological data. Smaller differences could complicate supporting or refuting the theories. For example, dogs have limited vocal modification abilities, but can learn to understand hundreds of human words. Thus some aspects of their auditory behavior and/or neurobiological substrates can be expected to be similar to how humans perceive speech, although dogs certainly lack, at least, the human capacity to comprehend spoken language. The motor theories for speech/song perception are also challenging to test since communication signals acoustically differ in a number of ways between animals and species as does the level of experience that animals have with species-typical communication vs. other sounds.

It is now well known that categorical perception is not unique to humans or to human speech as was originally thought (Ehret, 1987). Another potential challenge to this theory is that human and non-human primates appear to have comparable preferential responses for the processing of voice content in conspecific vocalizations (Belin et al., 2000; Petkov et al., 2008), some aspects of which have now been studied at the neuronal level in monkeys (Perrodin et al., 2011). If, however, the processing of voices that can be imitated in the human brain is subserved by processes that differ from the processes that support voice processing in monkeys, then such findings would likely support the motor theory of speech perception. If, however, these processes are shown to be largely comparable then the results might better support another theory, such as the *motor theory of vocal learning origin*.

Predictions of the motor theory of vocal learning origin

This theory underscores a distinction between the vocal *motor pathways* of complex-vocal learners and limited-vocal learners. Unlike the motor theories of speech/song perception, this theory makes no claims about whether the perceptual/learning systems differ between so-called vocal learners and non-learners. This is because the motor theory of vocal learning origin proposes a difference in the forebrain vocal motor pathway in vocal learners and non-learners (Feenders et al., 2008). Behaviorally, this theory is not mutually exclusive with the motor theories of song/speech perception; that is, the tenets of the latter theories can be interpreted to predict that vocal non-learners lack or have a limited access to the forebrain vocal motor pathway for perception. However, the motor theory of vocal learning origin does require that there are little to no differences in the auditory pathway input to the non-vocal motor pathway between vocal non-learners and learners.

Rigorously testing the motor theory of vocal learning origin will be challenging. Ideally one would use genetic/transgenic means to cause the forebrain auditory-motor pathway to duplicate during embryonic development and to form a direct projection from the forebrain to the brainstem motor neurons. Doing so would require discovering the genes that differ in their regulation of the vocal motor pathway or the adjacent non-vocal motor pathways. Candidate genes involved in axonal guidance and neuronal protection are being discovered in both song-learning birds and humans (Matsunaga and Okanoya, 2009; Hara et al., 2012; Horita et al., 2012). The impact on vocal behaviour from the genetic manipulations would need to be evaluated.

One potential challenge to parts of this theory is the observation that some complex-vocal learners (like humans and parrots) can synchronize their movements to a rhythmic beat in music (that is, to dance to a rhythm) whereas no vocal non-learners have been shown to be able to synchronize their movements in this way (Patel et al., 2009a,b; Schachner et al., 2009). The authors of these studies suggested that once the vocal learning pathway evolved, this affected the auditory pathway in such a way that it was differently connected with non-vocal motor pathways in complex-vocal learners. This hypothesis would predict differences in the auditory-motor pathway connectivity of vocal non-learners and learners. Another potential challenge to the motor theory of vocal learning origin is the finding in mice (Arriaga and Jarvis, in press; Arriaga et al., in press) of a limited forebrain vocal motor pathway with a sparse direct projection to brainstem vocal motor neurons, relative to complex-vocal learners. This would not negate the possibility that the vocal learning pathways emerged from a lineage of motor neurons related to the adjacent motor pathway, but it could mean that it might not be necessary to induce brain pathways by duplication. Instead, the vocal learning continuum hypothesis would predict that in complex vocal learners there was independent enhancement of an already existing pathway. In such a situation, the enhancement of the direct projections rather than their presence/absence may be the key difference between complex- and limited-vocal learners.

This theory might be tested with the use of viral vectors containing axonal guidance molecules to strengthen the sparse forebrain to brainstem vocal motor connectivity (e.g., cortico-bulbar projections). Positive outcomes from any such manipulations could be obtained by an animal being able to learn to more flexibly modify its vocalizations. A further possibility is that the innate brainstem vocal-production pathway may be able to separately support limited-vocal learning, such as, the pitch matching seen in mice and marmosets.

PREDICTIONS FROM GESTURAL THEORIES

Predictions of the gestural theory of spoken-language origin

The predictions of this theory are similar to the *motor theory of vocal learning origin* in that both predict a comparable perceptual system but differences in the production systems of complex-vocal learners. This theory goes a step further to hypothesize that brain pathways used to produce speech are intertwined with pathways used to perform learned gestures. Interestingly, Taglialatela and colleagues have observed Positron Emission Tomography (PET) activations in the chimpanzee inferior frontal cortex that occur after vocal production and gesturing but not after gesturing alone (Taglialatela et al., 2011). To more thoroughly test this theory, one would need to determine if the neurons and connectivity for gesturing (such as controlling hand movements) are the same as those for vocal production or form mixed neuronal subpopulations. Since, apes are a protected group, to study such neuronal populations might require developing a different non-human primate model system that can both gesture and vocalize, if possible. The gestural theory also relies on there being a forebrain motor cortical region that "controls" the sequencing of vocalizations and gestures in humans. Whether such a region would be found in the brains of non-human primates is uncertain. A critical test might require inactivating the currently poorly understood regions for vocal control and gesturing in certain non-human primates, and finding whether one set of behaviors can be maintained without the other. Efforts such as these could also help to clarify to what extent the motor theory of vocal learning origin depends on a gestural motor system.

Predictions of the gestural "mirror neuron" theory of language evolution

This theory predicts a lack of or weakness in the auditory-vocal motor link in limited-vocal learners or vocal non-learners. In other words, mirror neurons are engaged in human and non-human primates for gestural and other sensory-motor tasks but are not used in non-human primates for vocalization. One might predict that the mirror-neuron pathway for vocalization is: (1) not available for vocal production and imitation in limited-vocal learners such as chimpanzees, monkeys, many birds, etc.; and/or (2) that it is generally available for motor production in many animals but does not directly engage the auditory pathway or the pathway for innate vocal production. Given that to date linked auditory activated vocal mirror-neurons have only been reported in songbirds (Prather et al., 2008), a number of interesting issues remain to be tested across the species. For instance, are there auditory-vocal mirror-neurons engaged in vocal behavior in any of the limited-vocal learning birds, non-human primates or other vertebrates? The comparative connectivity data on the origins of the "arcuate fasciculus"—the classical language-related tract that links fronto-temporal brain regions in humans—remain controversial (Frey et al., 2008; Rilling et al., 2008) and do not seem able to currently provide strong evidence in support of or against a "weakness" in the auditory-vocal motor pathway in so-called vocal non-learners. Nevertheless, one way to functionally test whether the mirror neuron hypothesis is a viable mechanism for motor imitation would be to reversibly inactivate the so-called sensory-motor mirror neurons and determine if this affects vocal or other motor learning. This could be done with non-invasive trans-cranial magnetic stimulation in humans during speech processing, neuropharmacological inactivation in song-learning birds during song processing, or during some aspects of auditory/visual processing in any species with such neurons.

PREDICTIONS FROM A "SENSORIMOTOR INTEGRATION" MODEL

The "sensorimotor integration" model—of efference copies from prefrontal and premotor regions during speech production that modifies auditory processing and the perception of sounds—emphasizes the commonalities across the species regarding sensorimotor interactions (Rauschecker and Scott, 2009; Rauschecker, 2011). It thus differs in key ways from the cross species differences predicted by the motor theories of speech/song perception. This model could be largely compatible with the *motor theory of vocal learning origin* (section "Predictions of the Motor Theory of Vocal Learning Origin" above) if, as the motor theory proposes, the key difference between complex-vocal learners and limited-vocal learners is not in the perceptual system where sensori-motor interactions can help but in the form of the forebrain motor subsystem that is engaged. The model as proposed by Fee in songbirds is more applicable to sensorimotor pathways generally, suggesting that the song learning system has integrated with the adjacent motor pathways (Fee, 2012). Such a model is also consistent with the *motor theory of vocal learning origin*. Neuroimaging and neurophysiological data from multiple brain regions will be required to better evaluate the efference effects from frontal to auditory regions in complex- and limited-vocal learners. Results showing that the efference signal for vocal or other behavior differs in humans relative to non-human animals might better support the *motor theory of speech/song perception*. For further details on this model in humans, its historical basis and testable predictions see (Rauschecker and Scott, 2009; Rauschecker, 2011).

PREDICTIONS FROM COGNITIVE EVOLUTION HYPOTHESES: EVOLUTIONARY NEUROSCIENCE OF SYNTACTIC-RELATED PROCESSES

Cognitive hypotheses consider that complex-vocal learners, and more specifically humans, have enhanced capacities in cognitive systems broadly (e.g., learning, memory, attention, etc.) to support enhanced learned vocal communication perception or production. This is in contrast to notions that there are neurobiological substrates specifically dedicated to support speech/song capabilities. We discuss predictions of such hypotheses in four contexts: "domain" general and language specific predictions, and predictions revolving around primate and bird models.

General predictions from hypotheses on the evolution of cognitive systems

Behavioral predictions are that all animals can show varying levels of sensory or vocal learning, but are limited primarily by their cognitive abilities. Neurobiologically, limited-vocal learners would have the functionality of these systems with reduced capacity. The notion that most cognitive systems are improved in capacity in complex-vocal learners predicts a testable correlation between cognitive capacity and the level of engagement of cognitive processes (e.g., in learning increasingly more complex artificial-grammar sequence structures).

General support for cognitive evolution hypotheses are as follows. Some authors have obtained data that suggests that unlike humans, monkeys have reduced capacity for auditory recognition memory and may not directly engage the hippocampal memory circuit for auditory recognition memory (Fritz et al., 2005; Munoz-Lopez et al., 2010). In this regard, it remains to be determined whether human or songbird auditory recognition memory is indeed better than in non-human primates (or other limited-vocal learners), or if humans benefit from being able to semantically label speech to gain direct access to long-term memory circuits. The finding that only vocal learners can synchronize to a rhythm (Patel et al., 2009a) can support hypotheses on the evolution of cognitive systems. In this case, non-linguistic behavioral traits might be shown to depend on substrates that improved alongside or after the evolution of vocal learning. Additionally, there are various sorts of data, including anecdotal evidence, that some vocal learners (parrots, corvid songbirds, dolphins, elephants) have more complex cognitive behaviors relative to other animals that are more closely evolutionarily related to humans (Emery and Clayton, 2009). At a genetic level, it was recently discovered that humans have several unique duplications of the gene SRGAP2 (the SRGAP2 gene codes for SLIT-ROBO Rho GTPase-activating protein 2) that controls neural connectivity not found in any other primate or mammalian species tested to date (Charrier et al., 2012; Dennis et al., 2012); the extra copies when placed in mice induce an increase in dendritic spines and longer lasting spine immaturity, as is seen in human brains. The unique human gene duplications of SRGAP2 are hypothesized to be associated with greater learning capacity.

Additional comparative testing of behavioral abilities and of neurobiological, and genetic substrates is needed to provide stronger evidence either in favor of the evolution of cognitive "domain general" systems or alternatively in favor of "domain-specific" substrates that have considerably specialized for vocal production learning in humans or other vocal learners. Such testing will require two types of comparisons: (1) determining whether there are considerable vocal "domain specific" specializations not used for non-vocal learning capabilities (i.e., specializations in the auditory learning and vocal motor pathways); and (2) determining whether the auditory and vocal learning pathways in the brains of humans and other *vocal learners* function at a higher level of complexity than related brain pathways in other species. This would involve comparative analysis of cognitive and both auditory/vocal and non-auditory/non-vocal processing demands.

An interesting way forward for testing the tenets of various theories is with AGL and/or "statistical learning" paradigms, which are well suited for study in both humans and non-human animals. Neurobiological substrates for AGL and statistical learning can be evaluated in relation to language-related processes in the human brain. Such approaches seem well suited to test hypotheses on at least the evolution of cognitive systems and clarify the presence of domain general or domain specific substrates.

Syntactic complexity and the neurobiology of human language

Neuroimaging and neuropsychological work in humans has highlighted that how the human brain network for syntactic learning is engaged depends on the sequencing demands and types of structural relationships being evaluated (i.e., "syntactic complexity," Hagoort, 2009; Friederici, 2011; Fitch and Friederici, 2012; Petersson et al., 2012). From such results, a set of evolutionary "syntactic complexity" hypotheses have emerged (e.g., Friederici, 2004, 2011; Hurford, 2012; Petkov and Wilson, 2012). These propose that ancestral communication systems may have faced evolutionary pressures to manage greater sequencing demands in sensory input and/or motor output. This may have led to the evolution of enhanced systems in the human brain for processing syntactic complexity and the capacity for these to be learned during development (Friederici et al., 2011; Perani et al., 2011).

To bring this together in a model, **Figure 5A** schematizes the human auditory system engaged in syntactic perceptual learning, focusing, for brevity, on pathways interconnecting temporal and frontal cortical regions (for further details see, e.g., Friederici, 2011). Some of the model is based on the evidence in humans for brain regions and networks that are either sensitive to the violation of different types of learned AG structures (e.g., Friederici et al., 2006; Bahlmann et al., 2008, 2009; Friederici, 2009; Hagoort, 2009; Makuuchi et al., 2009; Petersson et al., 2012) and/or engaged in processing natural language structure (Marslen-Wilson and Tyler, 2007; Friederici, 2011; Tyler et al., 2011; Petersson et al., 2012). **Figure 5B** shows how the syntactic perceptual learning system is thought to engage with the vocal production learning system. These pathways are thought to explain different levels of complexity in AGL (Friederici, 2004; Friederici et al., 2006; Bahlmann et al., 2008; Friederici, 2011). For example, when humans build the initial syntactic structural analysis (such as evaluating only adjacent-relationships in a FSG), regions such as the frontal operculum (FOP) in the inferior prefrontal or insular cortex are engaged. The FOP interconnects with the anterior temporal lobe via the ventral uncinate fasciculus (UF) pathway (**Figure 5A**) (Friederici et al., 2006). However, when humans evaluate grammatical structures based on more distant non-adjacent relationships in FSG (Friederici, 2004; Petersson et al., 2012) or those having nested relationships (Bahlmann et al., 2008), then at least Brodmann area 44 (BA 44, a part of Broca's territory in the inferior frontal gyrus) is engaged in evaluating grammaticality (Friederici et al., 2006; Bahlmann et al., 2008; Folia et al., 2011; Petersson et al., 2012). BA44 is interconnected with the posterior temporal lobe via the dorsal superior-longitudinal fasciculus (SLF, **Figure 5B**), which includes parts of the human arcuate fasciculus. In humans, it is hypothesized that a different ventral (extreme capsule) pathway supports greater demands in syntactic complexity or syntactic-semantic relationships (Friederici, 2011). Unlike the ventral UF pathway interconnecting the FOP with the anterior temporal lobe, this other ventral pathway interconnects BA 45 and anterior temporal lobe regions (**Figure 5A** in orange, next to the FOP pathway in red) (Friederici, 2011). This model is similar to the dual pathways (dorsal and ventral streams) model for auditory and language processing, although there are also some differences (for reviews see: Rauschecker and Scott, 2009; Friederici, 2011; Rauschecker, 2011).

Non-human primate hypotheses: multiple pathways for "proto-syntactic" learning

Friederici (2004) proposed neurobiological substrates that in non-human primates might have been evolutionary substrates for proto-syntactic learning in humans. Petkov and Wilson (2012) extended this prediction into several subhypotheses, based on the finding that tamarin monkeys (Fitch and Hauser, 2004) and macaques (Wilson et al., 2011) can learn adjacent dependencies in an auditory AG with sequences that only require FSG processes. These are that: (1) the ventral, UF pathway (involving the monkey anatomical homologs of the FOP such as areas

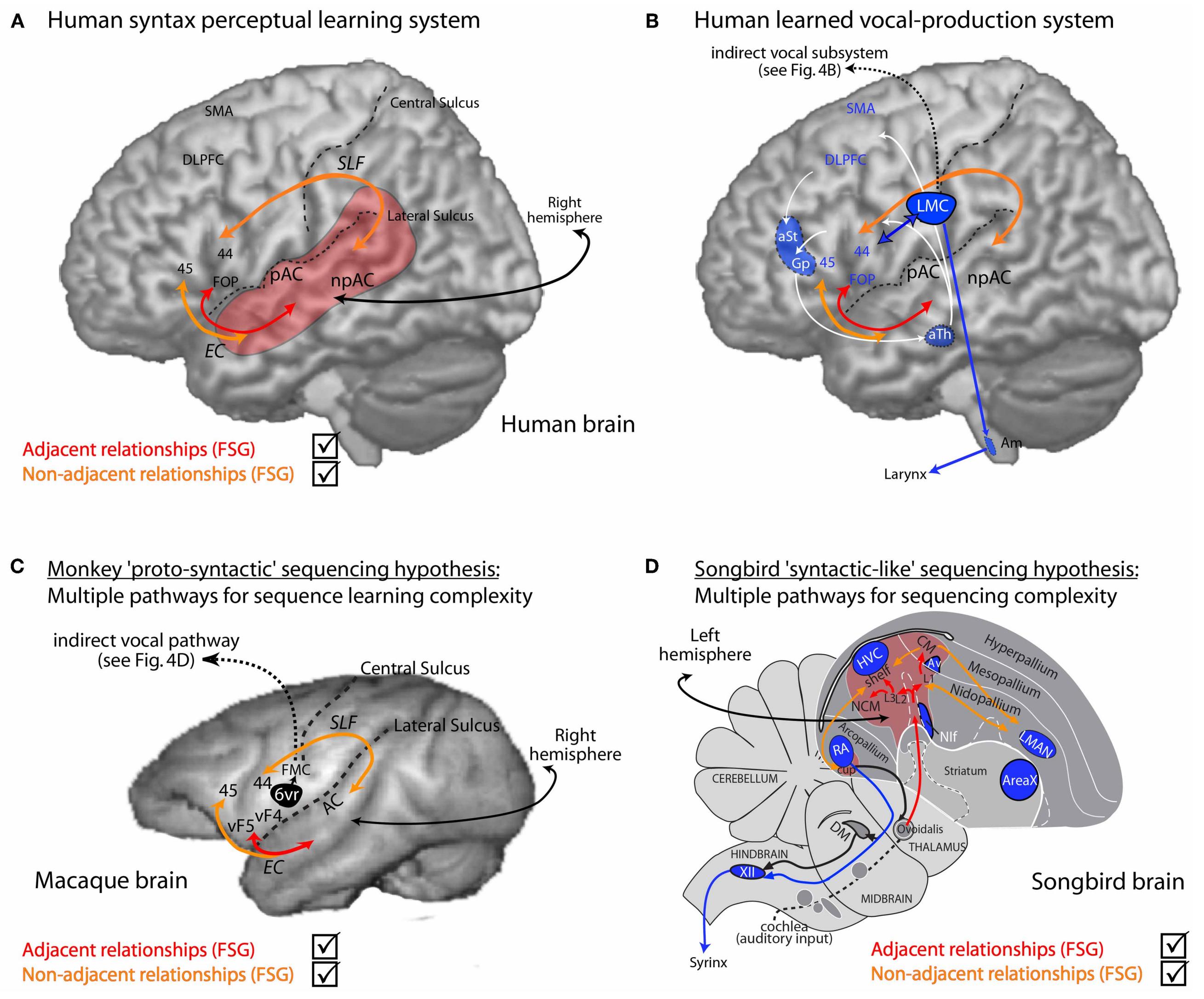


FIGURE 5 | Human syntactic learning and vocal production sub-systems, with hypothesized monkey and bird evolutionary substrates.
(A) Auditory perceptual learning system in humans (red and orange). Primary (pAC) and non-primary (npAC) auditory cortical regions are engaged in the auditory perceptual organization of sound. **(B)** The perceptual learning system interacts with a system for learned vocal production (blue, also see **Figure 4B**). **(C)** Hypothetical evolutionary "proto-syntactic" pathways that might be engaged in monkeys for the perceptual learning of different auditory sequence structures in Finite-State Artificial-Grammars (FSG), e.g., adjacent (red text) vs. non-adjacent (orange text) relationships (also see text). Note that the hypothetical ventral pathway is not expected to directly engage monkey Area 6vr (black) or the innate vocal production subsystem (black; see **Figure 4D**). More bilateral hemispheric engagement might be expected in non-human primates, see text, and/or that the cortical-striatal-thalamic loop would also be engaged in certain forms of implicit sequence learning. **(D)** Songbird auditory (red region and red/orange arrows) and song motor (blue regions) pathways. The auditory pathway is proposed to interact with motor regions adjacent to song nuclei for syntactic-like processing and production of vocal or non-vocal behaviors. Abbreviations: AC, auditory cortex; EC, extreme capsule fasciculus; SLF, superior-longitudinal fasciculus; UF, uncinate fasciculus. CM, caudal mesopallium; DLPFC, dorso-lateral prefrontal cortex; FMC, face motor cortex; FOP, frontal operculum; L2/L3, fields L2 and L3; NIf, interfacial nucleus of the nidopallium; NCM, caudal medial nidopallium; SMA, supplementary motor area; vF4/vF4, macaque anatomical regions ventral F4/F5; 44, 45, Brodmann Areas; See **Figure 4** for further abbreviations.

vF4/vF5 and parts of the inferior frontal insula) is engaged in the processing of adjacent relationships in AGs; and, (2) the dorsal, superior-longitudinal fasciculus pathway (including BA 44) is engaged for evaluating greater complexity in FSGs, such as non-adjacent relationships (Petersson et al., 2012), if the monkeys can learn these (**Figure 5C**). For further specifics and alternatives see: (Wilson et al., 2011).

Bird hypotheses: multiple pathways in songbirds for "syntactic-like" sensory learning

Regarding brain regions that might support songbird AGL, the study in Bengalese finches by Abe and Watanabe (2011) showed that expression of the immediate early gene *egr1* around the lateral magnocellular song nucleus of the anterior nidopallium (LMAN) was associated with whether these finches could learn aspects of AG sequences. However, it remains unclear whether these results relate to the AG structure or acoustical cues that were present in the "violation" sequences that were used for testing (see critique by: Berwick et al., 2011; Ten Cate and Okanoya, 2012). Another issue is that the areas studied in Bengalese finches, around LMAN, belong to the same nidopallial region that Feenders et al. (2008) found movement-driven (hopping) gene expression. Thus, it is not clear if the activation that was seen here could have resulted from differential movements of the animals to the different testing conditions or an association of the movement task with the grammatical processing. In all cases, we would hypothesize that some parts of the bird auditory system (in both complex-vocal learners and potentially also other birds) engages the motor areas adjacent to the song nuclei in evaluating AG structural relationships and to prepare non-vocal motor responses (**Figure 5D**). If so, a question that arises is in which neurobiological substrates would complex-vocal learners differ from more limited vocal learning birds or non-human primates? Possibly complex-vocal learners might be able to learn higher complexity in AG structures and for this potentially engage some parts of the vocal learning nuclei that would be unavailable to the limited vocal learners.

Relationship to predictions from motor and other theories

The hypotheses of **Figures 5C,D** do not illustrate the possible greater or lesser reliance on subcortical structures (such as the basal ganglia and thalamus) and/or cerebellum to support the learning of AG sequences. Some of these structures form a part of the system for motor-related learning and thus would link to and/or help to address predictions of motor theories. Moreover, how the animals learn AGs needs to be more carefully considered since we would expect different neurobiological substrates to be engaged if, for instance, the animals are engaged in implicit learning (e.g., habituated to grammatical sequences prior to testing) or are trained to discriminate grammatical vs. ungrammatical sequences, which would engage reward-dependent pathways (Petkov and Wilson, 2012). The predictions that we make in **Figure 5**, if supported could challenge the motor theory of speech perception/production that proposes considerable differences in the perceptual systems. Such results could instead support the motor theory of vocal learning origins. However, the hypotheses in this section cannot clarify whether the systems would depend on the gestural system, if no gestural or motor imitation component is involved in the experimental design. Moreover, the extent to which the systems in **Figure 5** differ across the species could also be used to test the sensori-motor integration or domain general hypotheses.

A major limitation in testing motor and other theories is that relatively much less effort has been made to study the basic behavioral phenotypes and underlying neural pathways that control either auditory-vocal or non-vocal pathways in vocal non-learning animals. As we have considered, such data can provide crucial insights on spoken language origins when compared with data from humans and complex-vocal learners, in which there has been a considerably greater focus. Thus, to validate or falsify the different hypotheses and to generate new ones, a much greater amount of additional comparative work is needed and any "one animal centric approaches" cannot be encouraged. We all tend to emphasize the literature and work in our own study groups or in a limited few species, but it remains important for researchers to continue to look beyond their immediate species of study.

CONCLUSIONS

This review has considered the behavioral and neurobiological data in complex-vocal learners such as, humans and songbirds and how they relate to data from so-called "vocal non-learners." We noted that the evidence provided by several recent examples in the animal behavioral literature motivates a revision of the hypothesized "vocal learning" vs. "vocal non-learning" distinction. We outlined an alternative hypothesis of greater variability in vocal learning categories and in a related but different behavioral phenotype, namely, auditory (sensory) sequence learning. Upon this modified perspective of the behavioral literature, we considered neurobiological distinctions between "vocal learners" and "non-learners," questioning whether these distinctions will be useful for clarifying behavioral results or whether behavioral variability can help us to understand the neurobiological substrates and distinctions at a finer level.

Motor, gestural and other theories were considered and predictions made, including from the perspective of animals that are not complex-vocal learners. We also considered the distinctions between the tenets of a number of theories regarding the sensory learning (reception) and motor (production) systems, including interconnectivity and interactions with the neurobiological systems supporting cognitive processes and how these may have evolved. This was used to make specific predictions, exemplified in humans, birds, and non-human primates to investigate the neurobiological substrates that might be evolutionarily related, either by common descent or convergence, to the ones that humans rely on for language.

Taking all of this into consideration, we highlighted the need for comparative approaches that more closely consider the behavioral and neurobiological data on sensory learning and vocal production abilities in many vertebrates. This will be critical for staying objective, empirically grounded, and realistic with regards to the aspects of human spoken language that different animal species could address or serve to model. We hope that this paper has been able to underscore the importance of and to encourage further interdisciplinary cross-species work for clarifying not only the origins of spoken language but also the conserved components and specializations present in the neurobiological systems supporting the communication abilities of animals.

ACKNOWLEDGMENTS

We are grateful to Colline Poirier, Kristina Simonyan and Benjamin Wilson for comments on previous versions of the paper. We also acknowledge the staff and students in our separate laboratories and Angela Friederici, William Marslen-Wilson, Kenny Smith, and Gustavo Arriaga for useful discussions. Financial support was provided by the Wellcome Trust (Christopher I. Petkov) and Howard Hughes Medical Institute (Erich D. Jarvis).

REFERENCES

Abe, K., and Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* 14, 1067–1074.

Allott, R. (1992). "The motor theory of language: origin and function," in *Language Origin: A Multidisciplinary Approach*, eds J. Wind, B. Chiarelli, B. Bichakjian, and A. Nocentini (Neatherlands: Kluwer Academic Publishers), 105–119.

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* 28, 105–124. discussion 125–167.

Arbib, M. A. (2010). Mirror system activity for action and language is embedded in the integration of dorsal and ventral pathways. *Brain Lang.* 112, 12–24.

Arnold, K., and Zuberbuhler, K. (2006). Language evolution: semantic combinations in primate calls. *Nature* 441, 303.

Arriaga, G., and Jarvis, E. D. (in press). Mouse vocal communication system: are ultrasounds learned or innate? *Brain Lang.* [Epub ahead of print].

Arriaga, G., Zhou, E., and Jarvis, E. D. (in press). Of mice, birds, and men: the mouse ultrasonic song system has features thought unique to humans and song learning birds. *PLoS ONE*.

Bahlmann, J., Schubotz, R. I., and Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *Neuroimage* 42, 525–534.

Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D., and Friederici, A. D. (2009). Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Res.* 1298, 161–170.

Baumann, S., Griffiths, T. D., Sun, L., Petkov, C. I., Thiele, A., and Rees, A. (2011). Orthogonal representation of sound dimensions in the primate midbrain. *Nat. Neurosci.* 14, 423–425.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312.

Berwick, R. C., Okanoya, K., Beckers, G. J., and Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* 15, 113–121.

Bickerton, D. (2009). "Syntax for Non-syntacticians: A brief primer," in *Biological Foundations and Origin of Syntax*, eds D. Bickerton and E. Szathmary. (Cambridge, MA: MIT Press), 3–15.

Bickerton, D., and Szathmary, E. (2009). *Biological Foundations and Origin of Syntax*. Cambridge, MA: MIT Press.

Bolhuis, J. J., Okanoya, K., and Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* 11, 747–759.

Bolhuis, J. J., and Wynne, C. D. (2009). Can evolution explain how minds work? *Nature* 458, 832–833.

Boughman, J. W. (1998). Vocal learning by greater spear-nosed bats. *Proc. Biol. Sci.* 265, 227–233.

Boumans, T., Gobes, S. M., Poirier, C., Theunissen, F. E., Vandersmissen, L., Pintjens, W., Verhoye, M., Bolhuis, J. J., and Van der Linden, A. (2008). Functional MRI of auditory responses in the zebra finch forebrain reveals a hierarchical organisation based on signal strength but not selectivity. *PLoS ONE* 3:e3184. doi: 10.1371/journal.pone.0003184

Briefer, E., and McElligott, A. G. (2011). Mutual mother-offspring vocal recognition in an ungulate hider species *(Capra hircus)*. *Anim. Cogn.* 14, 585–598.

Carr, C. E., and Code, R. A. (2000). "The central auditory system of reptiles and birds," in *Comparative Hearing: Birds and Reptiles*, eds R. R. Dooling, R. R. Fay, and A. N. Popper (New York, NY: Springer), 197–248.

Cartmill, M. (2010). "Primate classification and diversity," in *Primate Neuroethology*, eds A. A. Ghazanfar and M. L. Platt (Oxford: Oxford University Press), 10–30.

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

Charrier, C., Joshi, K., Coutinho-Budd, J., Kim, J. E., Lambert, N., De Marchena, J., Jin, W. L., Vanderhaeghen, P., Ghosh, A., Sassa, T., and Polleux, F. (2012). Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. *Cell* 149, 923–935.

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

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chong, T. T., Cunnington, R., Williams, M. A., Kanwisher, N., and Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Curr. Biol.* 18, 1576–1580.

Clarke, E., Reichard, U. H., and Zuberbuhler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS ONE* 1:e73. doi: 10.1371/journal.pone.0000073

Coude, G., Ferrari, P. F., Roda, F., Maranesi, M., Borelli, E., Veroni, V., Monti, F., Rozzi, S., and Fogassi, L. (2011). Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS ONE* 6:e26822. doi: 10.1371/journal.pone.0026822

Crockford, C., Herbinger, I., Vigilant, L., and Boesch, C. (2004). Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology* 110, 221–243.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. Princeton, NJ: Princeton University Press (1981 reprint of original).

Dennis, M. Y., Nuttle, X., Sudmant, P. H., Antonacci, F., Graves, T. A., Nefedov, M., Rosenfeld, J. A., Sajjadian, S., Malig, M., Kotkiewicz, H., Curry, C. J., Shafer, S., Shaffer, L. G., De Jong, P. J., Wilson, R. K., and Eichler, E. E. (2012). Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell* 149, 912–922.

DeWitt, I., and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* 109, E505–E514.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp. Brain. Res.* 91, 176–180.

Doupe, A. J., and Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* 22, 567–631.

Dubbeldam, J. L. (1998). The neural substrate for 'learned' and 'nonlearned' activities in birds: a discussion of the organization of bulbar reticular premotor systems with side-lights on the mammalian situation. *Acta anatomica* 163, 157–172.

Dujardin, E., and Jurgens, U. (2005). Afferents of vocalization-controlling periaqueductal regions in the squirrel monkey. *Brain Res.* 1034, 114–131.

Egnor, S. E., and Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends Neurosci.* 27, 649–654.

Ehret, G. (1987). "Categorical perception of sound signals: fact and hypotheses from animal studies," in *Categorical perception: The groundwork of cognition*, ed S. R. Harnad (Cambridge, UK: Cambridge University Press), 301–331.

Emery, N. J., and Clayton, N. S. (2009). Comparative social cognition. *Annu. Rev. Psychol.* 60, 87–113.

Esser, K. H. (1994). Audio-vocal learning in a non-human mammal: the lesser spear-nosed bat *Phyllostomus* discolor. *Neuroreport* 5, 1718–1720.

Evans, C. S., Evans, L., and Marler, P. (1993). On the meaning of alarm calls: functional reference in an avian vocal system. *Anim. Behav.* 46, 23–38.

Farries, M. A. (2004). The avian song system in comparative perspective. *Ann. N.Y. Acad. Sci.* 1016, 61–76.

Fee, M. S. (2012). Oculomotor learning revisited: a model of reinforcement learning in the basal ganglia incorporating an efference copy of motor actions. *Front. Neural Circuits* 6:38. doi: 10.3389/fncir.2012.00038

Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., Wada, K., Mouritsen, H., and Jarvis, E. D. (2008). Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS ONE* 3:e1768. doi: 10.1371/journal.pone.0001768

Fischer, J., and Hammerschmidt, K. (2011). Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. *Genes Brain Behav.* 10, 17–27.

Fitch, W. T., and Friederici, A. (2012). Artificial grammar learning meets formal language theory: an overview. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1933–1955.

Fitch, W. T., and Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.

Fitch, W. T., Huber, L., and Bugnyar, T. (2010). Social cognition and the evolution of language: constructing cognitive phylogenies. *Neuron* 65, 795–814.

Fitch, W. T., and Jarvis, E. D. (in press). "Birdsong and other animal models for human speech, song, and vocal learning," in *Language, Music and the Brain*, ed M. Arbib. (Cambridge MA: MIT Press).

Folia, V., Forkstam, C., Ingvar, M., Hagoort, P., and Petersson, K. M. (2011). Implicit artificial syntax processing: genes, preference, and bounded recursion. *Biolinguistics* 5, 105–132.

Frey, S., Campbell, J. S., Pike, G. B., and Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.* 28, 11435–11444.

Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends Cogn. Sci.* 8, 245–247.

Friederici, A. D. (2009). "Brain circuits of syntax," in *Biological Foundations and Origin of Syntax*, eds D. Bickerton and E. Szathmary. (Cambridge, MA: MIT Press), 239–252.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., and Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2458–2463.

Friederici, A. D., Mueller, J. L., and Oberecker, R. (2011). Precursors to natural grammar learning: preliminary evidence from 4-month-old infants. *PLoS ONE* 6:e17920. doi: 10.1371/journal.pone.0017920

Fritz, J., Mishkin, M., and Saunders, R. C. (2005). In search of an auditory engram. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9359–9364.

Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.

Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J., Gunnell, G., and Groves, C. P. (1998). Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* 9, 585–598.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23.

Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., Chojnowski, J. L., Cox, W. A., Han, K. L., Harshman, J., Huddleston, C. J., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Steadman, D. W., Witt, C. C., and Yuri, T. (2008). A phylogenomic study of birds reveals their evolutionary history. *Science* 320, 1763–1768.

Hage, S. R., and Jurgens, U. (2006). On the role of the pontine brainstem in vocal pattern generation: a telemetric single-unit recording study in the squirrel monkey. *J. Neurosci.* 26, 7105–7115.

Hagoort, P. (2009). "Reflections on the neurobiology of syntax," in *Biological Foundations and Origins of Syntax*, eds D. Bickerton and E. Szathmary (Cambridge, MA: MIT Press), 279–298.

Halpern, M. E., Gunturkun, O., Hopkins, W. D., and Rogers, L. J. (2005). Lateralization of the vertebrate brain: taking the side of model systems. *J. Neurosci.* 25, 10351–10357.

Hammerschmidt, K., Freudenstein, T., and Jurgens, U. (2001). Vocal development in squirrel monkeys. *Behaviour* 138, 1179–1204.

Hannig, S., and Jurgens, U. (2006). Projections of the ventrolateral pontine vocalization area in the squirrel monkey. *Exp. Brain. Res.* 169, 92–105.

Hara, E., Rivas, M. V., Ward, J. M., Okanoya, K., and Jarvis, E. D. (2012). Convergent differential regulation of parvalbumin in the brains of vocal learners. *PLoS ONE* 7:e29457. doi: 10.1371/journal.pone.0029457

Hast, M. H., Fischer, J. M., Wetzel, A. B., and Thompson, V. E. (1974). Cortical motor representation of the laryngeal muscles in *Macaca mulatta. Brain Res.* 73, 229–240.

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.

Heffner, H. E., and Heffner, R. S. (1984). Temporal lobe lesions and perception of species-specific vocalizations by macaques. *Science* 226, 75–76.

Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Curr. Anthropol.* 14, 5–24.

Holy, T. E., and Guo, Z. (2005). Ultrasonic songs of male mice. *PLoS Biol.* 3:e386. doi: 10.1371/journal.pbio.0030386

Honda, E., and Okanoya, K. (1999). Acoustical and syntactical comparisons between songs of the white-backed Munia (Lonchura striata) and its domesticated strain, the Bengalese finch (*Lonchura striata var. domestica*). *Zool. Sci.* 16, 319–326.

Hopkins, W. D., Taglialatela, J., and Leavens, D. A. (2007). chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Anim. Behav.* 73, 281–286.

Horita, H., Kobayashi, M., Liu, W.-C., Oka, K., Jarvis, E. D., and Wada, K. (2012). Specialized motor-driven dusp1 expression in the song system of separate lineages of vocal learning birds. *PLoS ONE* 7:e42173.

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

Hurford, J. R. (2012). *The Origins of Grammar: Language in the Light of Evolution*. Oxford: Oxford University Press.

Iwatsubo, T., Kuzuhara, S., Kanemitsu, A., Shimada, H., and Toyokura, Y. (1990). Corticofugal projections to the motor nuclei of the brainstem and spinal cord in humans. *Neurology* 40, 309–312.

Jaeger, G., and Rogers, J. (2012). Formal language theory: refining the Chomsky hierarchy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1956–1970.

Janik, V. M., and Slater, P. J. (1997). Vocal learning in mammals. *Adv. Study Behav.* 26, 59–99.

Janik, V. M., and Slater, P. J. (2000). The different roles of social learning in vocal communication. *Anim. Behav.* 60, 1–11.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Ann. N.Y. Acad. Sci.* 1016, 749–777.

Jarvis, E. D. (2006). Selection for and against vocal learning in birds and mammals. *Ornithol. Sci.* 5, 5–14.

Jarvis, E. D., Gunturkun, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D. J., Shimizu, T., Striedter, G., Wild, J. M., Ball, G. F., Dugas-Ford, J., Durand, S. E., Hough, G. E., Husband, S., Kubikova, L., Lee, D. W., Mello, C. V., Powers, A., Siang, C., Smulders, T. V., Wada, K., White, S. A., Yamamoto, K., Yu, J., Reiner, A., and Butler, A. B. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nat. Rev. Neurosci.* 6, 151–159.

Jarvis, E. D., and Mello, C. V. (2000). Molecular mapping of brain areas involved in parrot vocal communication. *J. Comp. Neurol.* 419, 1–31.

Jarvis, E. D., and Nottebohm, F. (1997). Motor-driven gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 94, 4097–4102.

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

Joly, O., Ramus, F., Pressnitzer, D., Vanduffel, W., and Orban, G. A. (2012). Interhemispheric differences in auditory processing revealed by fMRI in awake rhesus monkeys. *Cereb. Cortex* 22, 838–853.

Jurgens, U. (2002). Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* 26, 235–258.

Jurgens, U. (2009). The neural control of vocalization in mammals: a review. *J. Voice* 23, 1–10.

Jurgens, U., and Pratt, R. (1979a). The cingular vocalization pathway in the squirrel monkey. *Exp. Brain Res.* 34, 499–510.

Jurgens, U., and Pratt, R. (1979b). Role of the periaqueductal grey in vocal expression of emotion. *Brain Res.* 167, 367–378.

Kaas, J. H., and Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11793–11799.

Kaminski, J., Call, J., and Fischer, J. (2004). Word learning in a domestic dog: evidence for "fast mapping". *Science* 304, 1682–1683.

Khedr, E. M., and Aref, E. E. (2002). Electrophysiological study of vocal-fold mobility disorders using a magnetic stimulator. *Eur. J. Neurol.* 9, 259–267.

Kikusui, T., Nakanishi, K., Nakagawa, R., Nagasawa, M., Mogi, K., and Okanoya, K. (2011). Cross fostering experiments suggest that mice songs are innate. *PLoS ONE* 6:e17721. doi: 10.1371/journal.pone.0017721

Kirzinger, A., and Jurgens, U. (1982). Cortical lesion effects and vocalization in the squirrel monkey. *Brain Res.* 233, 299–315.

Kirzinger, A., and Jurgens, U. (1985). The effects of brainstem lesions on vocalization in the squirrel monkey. *Brain Res.* 358, 150–162.

Konishi, M., and Akutagawa, E. (1988). A critical period for estrogen action on neurons of the song control system in the zebra finch. *Proc. Natl. Acad. Sci. U.S.A.* 85, 7006–7007.

Kroodsma, D. E. (1976). Reproductive development in a female songbird: differential stimulation by quality of male song. *Science* 192, 574–575.

Kroodsma, D. E., and Konishi, M. (1991). A suboscine bird (*Eastern phoebe, Sayornis phoebe*) develops normal song without auditory feedback. *Anim. Behav.* 42, 477–487.

Kuypers, H. G. (1958a). Corticobular connexions to the pons and lower brain-stem in man: an anatomical study. *Brain* 81, 364–388.

Kuypers, H. G. (1958b). Some projections from the peri-central cortex to the pons and lower brain stem in monkey and chimpanzee. *J. Comp. Neurol.* 110, 221–255.

Kuypers, H. G. (1982). A new look at the organization of the motor system. *Prog. Brain Res.* 57, 381–403.

Lemon, R. N. (2008). Descending pathways in motor control. *Annu. Rev. Neurosci.* 31, 195–218.

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36.

Lilly, J. C. (1965). Vocal mimicry in tursiops: ability to match numbers and durations of human vocal bursts. *Science* 147, 300–301.

Lomber, S. G., and Malhotra, S. (2008). Double dissociation of 'what' and 'where' processing in auditory cortex. *Nat. Neurosci.* 11, 609–616.

Makuuchi, M., Bahlmann, J., Anwander, A., and Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8362–8367.

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

Margoliash, D., and Schmidt, M. F. (2010). Sleep, off-line processing, and vocal learning. *Brain Lang.* 115, 45–58.

Marler, P. (1970). Birdsong and speech development: could there be parallels? *Am. Sci.* 58, 669–673.

Marler, P. (2000). "Origins of music and speech: insights from animals," in *The Origins of Music*, eds N. L. Wallin, B. Merker and S. Brown (Cambridge: Massachusetts Institute of Technology), 31–50.

Marler, P., and Tamura, M. (1964). Culturally transmitted patterns of vocal behavior in sparrows. *Science* 146, 1483–1486.

Marshall, A. J., Wrangham, R. W., and Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees? *Anim. Behav.* 58, 825–830.

Marslen-Wilson, W. D., and Tyler, L. K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 823–836.

Masataka, N., and Fujita, I. (1989). Vocal learning of Japanese and rhesus monkeys. *Behaviour* 109, 191–199.

Matsunaga, E., and Okanoya, K. (2009). Vocal control area-related expression of neuropilin-1, plexin-A4, and the ligand semaphorin-3A has implications for the evolution of the avian vocal system. *Dev. Growth Differ.* 51, 45–54.

Milojevic, B., and Hast, M. H. (1964). Cortical motor centers of the laryngeal muscles in the cat and dog. *Ann. Otol. Rhinol. Laryngol.* 73, 979–988.

Miller, C. T., DiMauro, A., Pistorio, A., Hendry, S., and Wang, X. (2010). Vocalization induced cFos expression in marmoset cortex. *Front. Integr. Neurosci.* 4:128. doi: 10.3389/fnint.2010.00128

Mooney, R. (2009). Neurobiology of song learning. *Curr. Opin. Neurobiol.* 19, 654–660.

Moore, B. R. (2004). The evolution of learning. *Biol. Rev. Camb. Philos. Soc.* 79, 301–335.

Munoz-Lopez, M. M., Mohedano-Moriano, A., and Insausti, R. (2010). Anatomical pathways for auditory memory in primates. *Front. Neuroanat.* 4:129. doi: 10.3389/fnana.2010.00129

Murphy, R. A., Mondragon, E., and Murphy, V. A. (2008). Rule learning by rats. *Science* 319, 1849–1851.

Newport, E. L., Hauser, M. D., Spaepen, G., and Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cogn. Psychol.* 49, 85–117.

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., and Jenner, K. C. (2000). Cultural revolution in whale songs. *Nature* 408, 537.

Nottebohm, F. (1976). Vocal tract and brain: a search for evolutionary bottlenecks. *Ann. N.Y. Acad. Sci.* 280, 643–649.

Nottebohm, F. (1980). Brain pathways for vocal learning in birds: a review of the first ten years. *Prog. Psychobiol. Physiol. Psychol.* 9, 85–124.

Nottebohm, F., Stokes, T. M., and Leonard, C. M. (1976). Central control of song in the canary, *Serinus canarius*. *J. Comp. Neurol.* 165, 457–486.

Okanoya, K. (2004). The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. *Ann. N.Y. Acad. Sci.* 1016, 724–735.

Ouattara, K., Lemasson, A., and Zuberbuhler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proc. Natl. Acad. Sci. U.S.A.* 106, 22026–22031.

Owren, M. J., Dieter, J. A., Seyfarth, R. M., and Cheney, D. L. (1992). 'Food' calls produced by adult female rhesus (*Macaca mulatta*) and Japanese (*Macaca fuscata*) macaques, their normally-raised offspring, and offspring cross-fostered between species. *Behaviour* 120, 218–231.

Page, S. L., Chiu, C., and Goodman, M. (1999). Molecular phylogeny of Old World monkeys (Cercopithecidae) as inferred from gamma-globin DNA sequences. *Mol. Phylogenet. Evol.* 13, 348–359.

Pallier, C., Devauchelle, A. D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2522–2527.

Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (2009a). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19, 827–830.

Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (2009b). Studying synchronization to a musical beat in nonhuman animals. *Ann. N.Y. Acad. Sci.* 1169, 459–469.

Penfield, W., and Rasmussen, T. (1949). Vocalization and arrest of speech. *Arch. Neurol. Psychiatry* 61, 21–27.

Pepperberg, I. M. (2010). Vocal learning in Grey parrots: a brief review of perception, production, and cross-species comparisons. *Brain Lang.* 115, 81–91.

Perani, D., Saccuman, M. C., Scifo, P., Awander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., and Friederici, A. D. (2011). Neural language networks at birth. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16056–16061.

Perrodin, C., Kayser, C., Logothetis, N. K., and Petkov, C. I. (2011). Voice cells in the primate temporal lobe. *Curr. Biol.* 21, 1408–1415.

Perruchet, P., and Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychon. Bull. Rev.* 12, 307–313.

Petersson, K. M., Folia, V., and Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain Lang.* 120, 83–95.

Petkov, C. I., Kayser, C., Augath, M., and Logothetis, N. K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol.* 4:e215. doi: 10.1371/journal.pbio.0040215

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (2008). A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–374.

Petkov, C. I., Logothetis, N. K., and Obleser, J. (2009). Where are the human speech and voice regions and do other animals have anything like them? *Neuroscientist* 15, 419–429.

Petkov, C. I., and Wilson, B. (2012). On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2077–2088.

Phan, M. L., and Vicario, D. S. (2010). Hemispheric differences in processing of vocalizations depend on early experience. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2301–2306.

Pinker, S. (1994). *The Language Instinct*. New York, NY: William Morrow and Company, Inc.

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

Poirier, C., Boumans, T., Verhoye, M., Balthazart, J., and Van Der Linden, A. (2009). Own-song recognition in the songbird auditory pathway: selectivity and lateralization. *J. Neurosci.* 29, 2252–2258.

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

Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P., and Mishkin, M. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature* 427, 448–451.

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

Ralls, K., Fiorelli, P., and Gish, S. (1985). Vocalizations and vocal mimicry in captive harmor seals, *Phoca vitulina*. *Can. J. Zool.* 63, 1050–1056.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Curr. Opin. Neurobiol.* 8, 516–521.

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear. Res.* 271, 16–25.

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Front. Evol. Neurosci.* 4:7. doi: 10.3389/fnevo.2012.00007

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.

Reber, A. S. (1967). Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* 6, 855–863.

Rey, A., Perruchet, P., and Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: evidence from baboons (*Papio papio*). *Cognition* 123, 180–184.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., and Behrens, T. E. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.

Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136.

Rumbaugh, D. M. (1977). *Language Learning by a Chimpanzee*. New York, NY: Academic Press.

Saffran, J. R., Aslin, 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.

Saranathan, V., Hamilton, D., Powell, G. V., Kroodsma, D. E., and Prum, R. O. (2007). Genetic evidence supports song learning in the three-wattled bellbird Procnias tricarunculata (Cotingidae). *Mol. Ecol.* 16, 3689–3702.

Schachner, A., Brady, T. F., Pepperberg, I. M., and Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr. Biol.* 19, 831–836.

Schmitt, D. (2010). "Primate locomotor evolution: Biomechanical studies of primate locomotion and their implications for understanding primate neuroethology," in *Primate Neuroethology*, eds A. A. Ghazanfar and M. L. Platt (Oxford University Press), 10–30.

Schusterman, R. J. (2008). "Vocal learning in mammals with special emphasis on pinnipeds," in *The Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication*, eds D. K. Oller and U. Griebel (Cambridge, MA: MIT Press).

Scott, S. K., McGettigan, C., and Eisner, F. (2009). A little more conversation, a little less action–candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.* 10, 295–302.

Seyfarth, R. M., and Cheney, D. L. (1986). Vocal development in vervet monkeys. *Anim. Behav.* 34, 1640–1658.

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 210, 801–803.

Shettleworth, S. J. (2010). *Cognition, Evolution, and Behavior*. Oxford: Oxford University Press.

Sibley, C. G., and Ahlquist, J. E. (1990). *Phylogeny and Classification of Birds*. New Haven, Connecticut: Yale University Press.

Simonyan, K., and Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *Neuroscientist* 17, 197–208.

Simonyan, K., and Jurgens, U. (2003). Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Res.* 974, 43–59.

Simões, C. S., Vianney, P. V. R., de Moura, M. M., Freire, M. A. M., Mello, L. E., Sameshima, K., Araújo, J. F., Nicolelis, M. A. L., Mello, C. V., and Ribeiro, S. (2010). Activation of frontal neocortical areas by vocal production in marmosets. *Front. Integr. Neurosci.* 4:123 doi:10.3389/fnint.2010.00123

Simpson, H. B., and Vicario, D. S. (1990). Brain pathways for learned and unlearned vocalizations differ in zebra finches. *J. Neurosci.* 10, 1541–1556.

Snowdon, C. T. (2009). "Plasticity of communication in nonhuman primates," in *Advances in the Study of Behavior*, eds M. Naguib and V. M. Janik (Burlington, NJ: Academic Press), 239–276.

Sugiura, H. (1998). Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Anim. Behav.* 55, 673–687.

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

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., and Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PLoS ONE* 6:e18852. doi: 10.1371/journal.pone.0018852

Tallerman, M. (2011). "What is syntax?," in *The Oxford Handbook of Language Evolution*, eds M. Tallerman and K. R. Gibson (Oxford: Oxford University Press), 442–455.

Ten Cate, C., and Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1984–1994.

Teufel, C., Ghazanfar, A. A., and Fischer, J. (2010). On the relationship between lateralized brain function and orienting asymmetries. *Behav. Neurosci.* 124, 437–445.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290–293.

Tomasello, M., Savage-Rumbaugh, S., and Kruger, A. C. (1993). Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Dev.* 64, 1688–1705.

Troyer, T. W., and Doupe, A. J. (2000). An associational model of birdsong sensorimotor learning I. Efference copy and the learning of song syllables. *J. Neurophysiol.* 84, 1204–1223.

Turing, A. (1968). "Intelligent machinery," in *Cybernetics: Key Papers*, eds C. R. Evans and A. D. J. Robertson (Baltimore, MD: University Park Press).

Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., Papoutsi, M., and Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain* 134, 415–431.

Uddén, J., Ingvar, M., Hagoort, P., and Petersson, K. M. (2012). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: investigating the push-down stack model. *Cogn. Sci.* doi: 10.1111/j.1551-6709.2012.01235.x. [Epub ahead of print].

van Heijningen, C. 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.

Vates, G. E., Broome, B. M., Mello, C. V., and Nottebohm, F. (1996). Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *J. Comp. Neurol.* 366, 613–642.

Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., and Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates* 50, 56–64.

Wild, J. M. (1997). Neural pathways for the control of birdsong production. *J. Neurobiol.* 33, 653–670.

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

Wilson, B., Collison, M. G., Slater, H., Smith, K., Marslen-Wilson, W., and Petkov, C. (2011). Behaviour and functional imaging of 'artificial-grammar' sequence learning in rhesus macaques. *Soc. Neurosci. Abstr.* Program No: 173.18 (p. 53).

Yip, Z. C., Miller-Sims, V. C., and Bottjer, S. W. (2012). Morphology of axonal projections from the high vocal center to vocal motor cortex in songbirds. *J. Comp. Neurol.* 520, 2742–2756.

Zimmerman, E. (1989). Aspects of reproduction and behavioral and vocal development in Senegal bushbabies (*Galago senegalensis*). *Int. J. Primatol.* 10, 1–16.

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: 01 December 2011; accepted: 20 July 2012; published online: 16 August 2012.

Citation: Petkov CI and Jarvis ED (2012) Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. Front. Evol. Neurosci. ***4****:12. doi: 10.3389/fnevo.2012.00012*