

UC Irvine

UC Irvine Previously Published Works

Title

Neurobiological roots of language in primate audition: common computational properties

Permalink

<https://escholarship.org/uc/item/60c2f4wd>

Journal

Trends in Cognitive Sciences, 19(3)

ISSN

1364-6613

Authors

Bornkessel-Schlesewsky, Ina
Schlesewsky, Matthias
Small, Steven L
[et al.](#)

Publication Date

2015-03-01

DOI

10.1016/j.tics.2014.12.008

Peer reviewed



Published in final edited form as:

Trends Cogn Sci. 2015 March ; 19(3): 142–150. doi:10.1016/j.tics.2014.12.008.

Neurobiological roots of language in primate audition: common computational properties

Ina Bornkessel-Schlesewsky^{1,2}, Matthias Schlesewsky³, Steven L. Small⁴, and Josef P. Rauschecker^{5,6}

¹Cognitive Neuroscience Laboratory, School of Psychology, Social Work and Social Policy, University of South Australia, Adelaide, Australia

²Department of Germanic Linguistics, University of Marburg, Marburg, Germany

³Department of English and Linguistics, Johannes Gutenberg-University, Mainz, Germany

⁴Department of Neurology, University of California, Irvine, CA, USA

⁵Laboratory of Integrative Neuroscience and Cognition, Department of Neuroscience, Georgetown University Medical Center, Washington D.C., USA

⁶Institute for Advanced Study, Technische Universität München, Garching, Germany

Abstract

This paper presents a new perspective on an old question: how does the neurobiology of human language relate to brain systems in nonhuman primates? We argue that higher-order language combinatorics – including sentence and discourse processing – can be situated in a unified, cross-species dorsal-ventral streams architecture for higher auditory processing, and that the functions of the dorsal and ventral streams in higher-order language processing can be grounded in their respective computational properties in primate audition. This view challenges an assumption, common in the cognitive sciences, that a nonhuman primate model forms an inherently inadequate basis for modeling higher-level language functions.

How does the brain implement language?

Neurobiologically plausible models of human brain function are typically based on detailed animal models. However, while the applicability of this modeling strategy is widely accepted for domains such as vision or audition, its transferability to human language is considerably more controversial. The reason for this perspective – particularly at the level of sentences and above – relates to complex computational properties of human grammars and their purported specificity to our species [1,2].

© 2014 Elsevier Ltd. All rights reserved.

Corresponding Author: Ina Bornkessel-Schlesewsky (Ina.Bornkessel-Schlesewsky@unisa.edu.au).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

With respect to neurobiological models of speech and language, these considerations have led to an interesting dualism. It is generally accepted that human speech and language processing is supported by a cortical dorsal-ventral-streams architecture that shares many anatomical characteristics with the extended auditory system of nonhuman primates (e.g. [3–8]). This architecture involves a division of labor between two cortical streams of information transfer from auditory cortex (AC) to prefrontal regions. As shown in more detail in Figure 1, the postero-dorsal stream connects AC to the posterior and dorsal part of inferior frontal cortex (IFC) (Brodmann area [BA] 44) via posterior superior temporal (pST) cortex, inferior parietal lobule (IPL), and premotor cortex (PMC); the antero-ventral stream, by contrast, traverses anterior superior temporal cortex (aST) to terminate in more anterior and ventral parts of the inferior frontal gyrus (BA 45). Importantly, most models in this domain have focused primarily on speech and word processing, rather than on the complex combinatorial properties of language claimed to be unique to humans. The few available dual-stream models of sentence processing, by contrast, typically assume that the neural circuitry of nonhuman primates is insufficient to support sentence comprehension because of a fundamental difference in its computational architecture that is not simply a matter of degree (e.g. [8]). They thus posit uniquely human additions to this circuitry in the dorsal stream, which are assumed to have evolved late from a phylogenetic perspective and to mature late from an ontogenetic perspective [9]. Hence, in spite of the broad consensus regarding the anatomical overlap between the primate auditory system and the cortical speech and language architecture, it is typically assumed that the nonhuman primate system is neither quantitatively nor qualitatively sufficient to support the computational needs of higher-order language (i.e., sentence and discourse) processing.

In addition, recent research has even questioned the necessity of a neural architecture akin to that of the primate auditory system for the computational mechanisms underlying higher-order language. As nonhuman primates are generally considered to not be complex vocal learners, there has been an increased interest in alternative animal models, focusing on species that do show vocal learning abilities. In this context, songbirds have played a dominant role, based on the shared ability for complex sequence processing in avians and humans (e.g., [10,11]). Thus, by shifting the focus onto evolutionary convergence as opposed to common descent, birdsong models have further perpetuated the move away from a nonhuman primate model for the neurobiology of higher-order language [2,10] – the importance of such a model for basic aspects of speech, and possibly word-level processing notwithstanding. (For approaches advocating the comparison of multiple nonhuman animal models, see e.g. [12,13].)

Here, we argue that the tendency to abandon the nonhuman primate auditory system as a suitable animal model for the neurobiology of higher-order language may be premature. (For a similar, recent argument regarding the evolution of speech, see [14].) To the contrary, we suggest that, when the computational requirements for sentence and discourse processing are broken down into more basic mechanistic components, there is indeed quite compelling evidence to suggest that the computational architecture of the nonhuman primate dorsal and ventral auditory streams is qualitatively sufficient for performing the requisite computations. In other words, the basic computational building blocks necessary for language processing are already in place in the nonhuman primate, though the system lacks the necessary

quantitative scale to support language. We also offer some suggestions as to why the primate auditory system may in fact even fulfill the stronger criterion of necessity, based on the notion of cross-stream integration and the role of prefrontal cortex.

Basic computational properties underlying higher-order language combinatorics

One of the hallmarks of higher-order language (sentences and discourse) in humans is its combinatorial flexibility, i.e., the ability to combine smaller units into larger units in order to express a wide range of meanings. Two basic combinatorial mechanisms are generally agreed upon in linguistic theory: (i) the combination of elements into sequences (i.e., combining elements A and B to form the sequence A-before-B); (ii) the combination of elements to form dependencies, independent of sequential order. To illustrate the difference between the two mechanisms, consider the phrase "the red boat" and its French counterpart "le bateau rouge" (literally: 'the boat red'). In both cases, red describes a property of the boat (i.e., there is a dependency between boat and red), but the sequential order in which the two words are expressed differs between the two languages. How strongly dependencies rely on particular sequential orders also differs across the languages of the world. While in English, sequential order is the primary cue for extracting dependencies (e.g., "The boy kissed the girl" can only mean that the boy was the kisser and the girl the person being kissed), most other languages show more flexibility in this regard (e.g., German, where "Den Jungen küsste das Mädchen", 'The boy.ACCUSATIVE kissed the girl.NOMINATIVE', means that the girl kissed the boy; with dependencies indicated here by the accusative and nominative case marking). As is apparent from these simple examples, sequencing and dependency formation are basic – and separable – properties of human language, and all theories of grammar assume these two computational mechanisms in one form or another (e.g., [15–18]).

We propose here that sequence-based (order-sensitive) and dependency-based (order-insensitive) combinatorics are supported by the dorsal and ventral cortical streams, respectively. Box 1 discusses this assumption in more detail in the context of neurobiologically motivated design principles. In the following sections, we review evidence for this claim from language studies and demonstrate that these basic computational functions are already in place in the dual auditory streams of nonhuman primates.

Antero-ventral stream: Computation of successively more complex auditory objects

In nonhuman primates, the ventral stream subserves the recognition of successively more complex auditory objects, ranging from elementary auditory features (e.g., frequency-modulated (FM) sweeps or bandpass noise bursts) to species-specific vocalizations (monkey calls) [4,19,20]. Auditory object formation is a form of categorization, in which spectro-temporal properties are grouped into perceptual [21] and, at higher hierarchical levels,

conceptual units. It thus provides the computational basis for an elementary mapping from spectro-temporal patterns to concepts.

Language also involves the identification of auditory objects (e.g., phonemes, syllables, words, phrases), with evidence for a similar hierarchical organization of auditory object recognition along the antero-ventral stream [22]. This involves the mapping from spectro-temporal patterns to sensory-independent categories (e.g., [23] for syllables) and time-invariant semantic structures (concepts) [5–7]. Notably, time-invariance constitutes a crucial property of conceptual representations (see Box 1). Further evidence to implicate the ventral stream in conceptual processing in humans stems from the deficits shown by patients with Semantic Variant Primary Progressive Aphasia (semantic PPA; also known as semantic dementia, though the two terms are not completely synonymous [24]). This condition is characterized by atrophy of the anterior temporal lobes (e.g., [25,26]). Accordingly, many researchers view portions of the anterior temporal lobe as a hub of semantic processing, binding together perceptually-based semantic representations into coherent concepts (but see [27,28]). Converging evidence for this view stems from various experimental approaches, including neuroimaging [7,29], transcranial magnetic stimulation [30,31], and computational modeling [6].

Although the involvement of the ventral stream in mapping auditory input onto conceptual representations is widely accepted, it has been considerably less clear to date how this perspective on ventral stream function might be integrated with the results of a separate line of research, which suggests that the anterior temporal lobe also contributes to combinatorial processing [22,32,33], an assumption that features in several dual-stream models of speech and language [5,8].

We propose (see also [34]) that these two functions of the ventral stream – auditory-conceptual mappings and combinatorial processing – can be subsumed under a single mechanism, which is moreover motivated by independent assumptions regarding the recognition of auditory objects and hierarchical processing (see Box 1). Auditory objects can be rendered more complex (feature-rich) by combining an increasing number of attributes. An important characteristic of this combinatorial property is its commutativity, i.e., order is not important ($AB = BA$). In other words, adding information leads to an updating of existing auditory objects via the formation of dependencies rather than the formation of a sequence of auditory objects, as in the dorsal stream (see below). For example, the auditory object "boat" is modified to include an additional property (that it is red) via processing of the word "red", irrespective of whether "red" occurs before "boat" or after. Like auditory objects, dependencies are formed based on grouping cues (e.g., case marking, likelihood of co-occurrence).

Postero-dorsal stream: Sequential processing via an internal model

In keeping with its visual counterpart, the postero-dorsal auditory stream was originally characterized as a “where” [35] or “how” [36] stream [3,20]. Besides processing auditory space and motion, it subserves auditory-motor mappings [4,37,38], including the processing of sound sequences (e.g., [39–41]) and speech rhythm [42]. The computational machinery

common to these functions can be characterized in terms of an internal (forward) model [19], which serves to predict upcoming sensory events within a sequence on the basis of the previous input [43–48].

For sentence comprehension, there is substantial evidence linking the dorsal stream to the processing of word order (i.e., sentence-level sequencing). A number of fMRI studies comparing sentences with object-before-subject and subject-before-object orders have reported activation in regions associated with the dorsal stream, including pST, IPL, and PMC regions, as well as the IFG [49–51]. Crucially, findings such as these do not implicate the dorsal stream as a syntax stream, since the activations in question also rely on non-syntactic (e.g., semantic) parameters ([50,51]).

As proposed for the primate model and for speech processing [4,19], the implementation of a forward model provides a unified functional explanation for the dorsal-stream sequence processing capability at different linguistic levels. To explain the interplay among the various, hierarchically organized levels of sequences (discourse is composed of sentences which are made up of words which comprise sounds), a hierarchically organized set of forward models is required, each yielding predictions for the next level down (see Box 1) [46,52]. Similar claims have been made in the domains of action control, which shows a similar hierarchical organization [53,54], and of vision [55].

Evidence for the computational division of labor between streams

Experimental evidence from a variety of different methods provides direct support for the computational division of labor between streams as proposed here. One study used fMRI to demonstrate a parametric increase in activation of both anterior (ventral stream) and posterior (dorsal stream) temporal lobes, the latter in the posterior superior temporal sulcus (pSTS), in response to linguistic constituents of increasing size [56]. However, the increase proved to be dependent on the use of real words in the anterior but not posterior temporal lobe. Posteriorly, the pSTS showed a similar parametric activation increase for sequences of real words and phonologically legal nonwords with grammatical features preserved (i.e., pseudowords). This attests to the importance of conceptual combinatorics for the ventral stream, while the dorsal stream responds to more abstract sequencing demands.

Neuroimaging studies with PPA patients confirm this assumption. Syntactic deficits in PPA patients – evidenced by a reduced ability to comprehend sentences not adhering to the typical (canonical) Agent-Action-Object sequence for English [57] – correlate with damage to dorsal, but not ventral white-matter tracts [58]. Conversely, semantic PPA patients with atrophy of the anterior temporal lobe show comparable dorsal stream fMRI activation to control subjects for non-canonical sentences [59]. This finding suggests that the known increase in activation of the anterior temporal lobe for sentences versus word lists (see [32] for a review) should be attributed to a different combinatorial mechanism than that engendering dorsal-stream activation for non-canonical sentences, i.e., building commutative (order-insensitive) dependencies rather than processing sequences.

Cross-stream integration and prefrontal cortex

In the model proposed here, prefrontal cortex is viewed as a controller of information flow along the two streams [60], subserving cross-stream integration and mediating top-down feedback from one stream to the other [61,62]. From this perspective, integration can be envisaged as a recoding of information originating in one stream into a signal for top-down modulation of the other (see also [4]). (Note, however, that in addition to cross-stream integration via feedback loops, there is some evidence for interactive cross-talk between streams [61] and for subcortical contributions – particularly striatal – to the neural language architecture [63]).

In contrast to several other current neurobiological models of language [8,64], our proposal does not imply a specific involvement of prefrontal cortex – and particularly the IFG – in language processing *per se* (e.g., in accomplishing unification or syntactic processing). Rather, it subscribes to the view that the prefrontal cortex performs domain-general regulatory functions that play a nonspecific but crucial role in language processing [65–67]. (For a review of recent evidence from the language domain, see [34]). Increased regulation (or cognitive control) becomes necessary when there is a need for top-down biases to mediate among conflicting representations or when the recoding of information for cross-stream interaction leads to an incompatibility with the current state of the stream on which the top-down influence is exerted. A prediction following from this assumption is that the prefrontal cortex should be capable of up- or down-regulating activity within a particular stream in accordance with current processing demands.

On sufficiency: Differences between nonhuman primates and humans

So far, the focus has been on commonalities between humans and nonhuman primates regarding the computational properties of the dorsal and ventral auditory streams. Nevertheless, nonhuman primates obviously do not have language – at least not in a comparable way to humans – and they do not appear to be able to acquire human language even when taught in gestural form [68]. How can a proposal positing only quantitative neurocomputational differences between humans and nonhuman primates – and, hence, a qualitative sufficiency for higher-order language processing of the computational mechanisms supported by the auditory system of nonhuman primates – account for such differences? Importantly, in keeping with the approach of examining the basic computational building blocks underlying higher-order language, we focus on the capacity of nonhuman primates for information processing rather than on their production abilities, which may depend on constraints of a fundamentally different nature (see [69]; for a similar argument, see [12]).

With regard to the ventral stream, the potential complexity (feature-richness) of auditory objects in humans is undoubtedly higher than in nonhuman primates. This may, in part, be due to the possibility of auditory object combination (or unification in linguistic theory), as outlined above. Nevertheless, whether the capacity for auditory object combination is unique to humans or whether nonhuman primates also show certain limited capacities in this regard remains to be investigated. Certainly, the ability to combine features at the cellular level

(combination sensitivity) is well documented for early stages of auditory cortex in nonhuman primates as well as other animals [70,71], rendering neurons selective, for example, to monkey calls as opposed to more elemental features such as FM sweeps or band-passed noise bursts [72]. It is not clear to date whether this also applies to higher levels of processing.

However, it is the dorsal stream that has been primarily implicated in the unique computational capacities assumed to support human language [2,8,9,73]. In contrast to these approaches, we do not subscribe to the notion (see Box 2) that, beyond basic sequence processing, a more elaborate and qualitatively distinct computational mechanism (i.e., discrete infinity produced by recursion) is required for human language. We assume instead, as proposed in [34], that the ability to combine two elements A and B in an order-sensitive manner to yield the sequence AB forms the computational basis for the processing capacity of the postero-dorsal stream in human language. Indeed, nonhuman primates show at least a very basic capacity for this type of simple sequence processing, e.g., chimpanzees can learn to produce ordered 2-word sequences [68] or imitate sequentially structured actions [74]. Nonhuman primates are also proficient at transitive inference (TI), i.e., they can correctly infer from the sequential presentation of paired stimuli that, if A precedes B and B precedes C, A also precedes C. In TI tasks, rhesus monkeys show similar performance in terms of error rates and reaction times to 4- and 6-year-old children [75,76]. Findings such as these, as well as results on numerical reasoning [77], have been used to argue for a spatial representation of serial order information [78], a suggestion that resonates with the function of the dorsal stream in both spatial and sequence processing.

Furthermore, neuroanatomical tracer studies in monkeys provide evidence for direct connections from posterior-dorsal auditory regions to parietal and prefrontal regions [79–81], similar to those in humans [82], which can form the substrate for auditory-motor coupling required for these forms of neurocomputation. These existing structures clearly enable demonstrated audio-motor behavior in monkeys [83,84]. Thus, while monkeys may not possess the level of sophistication in this domain enabling them to have speech and complex vocal learning, they certainly have the machinery necessary to link perception and action. Such sensorimotor control loops, commonly referred to as internal models, provide the basis for sound sequences being matched to action sequences and we posit here that the basic mechanism of sequence processing, which we claim is subserved by the dorsal stream, relies on the computations afforded by an internal model (see Box 1). Crucially, our focus is on sequencing as a basic computational mechanism in higher-order language rather than in speech processing. In other words, we are not claiming that monkeys have either articulatory or acoustic phonetics, but that they have brain structures that could support – given other factors – the essential structures of higher-order language. While the lack of a human-like articulatory apparatus could lead to evolutionary structural differences in the brain mechanisms for phonetics, we are not aware of any evidence that such differences need necessarily impact higher order combinatorial processing.

We posit that the difference in sequence processing capacity between humans and nonhuman primates – and hence the considerably more complex structure of human language – can be traced back to the hierarchical system of forward models implemented by

the human dorsal stream. By contrast, it appears likely that the complexity of the forward model implemented by the dorsal stream of nonhuman primates is hierarchically limited (e.g., by a smaller number of specialized cortical areas [85], smaller cortical volume [86], or smaller number of connecting axons), thus accounting for the limitations shown by nonhuman primates in the learning of hierarchically structured sequential action sequences (i.e., the reduced hierarchical “depth” shown in comparison to humans [87]).

Thus, it appears that the limitations of nonhuman primate vocal productions and their vocal learning abilities may not be key to understanding the possible neurobiological roots of higher-order language in the nonhuman primate brain. In further support of this assumption, a separate line of research has identified possible precursors to human speech and language in the rhythm of nonhuman primate oro-facial movements [14,88]. Recent studies in rhesus macaques suggest that lip smacking behavior may be particularly relevant in this regard, as it shows similar rhythmic properties to human speech and language, namely a rhythmical structure in the 3–8 Hz range [89,90]. It further shares a range of additional characteristics with speech, including similar developmental trajectories of rhythmical structure in production as well as perceptual tuning to the preferred frequency range in perception ([91], for a review). Finally, recent findings show that new-world monkeys (marmosets) engage in turn-taking during their vocal communicative behaviour and, in doing so, manifest multiple signs of cooperative coordination, including waiting before responding to prevent overlapping calls, call coupling and reciprocal entrainment of calls between monkeys [92]. This suggests that some of the fundamental prerequisites for discourse communication are already in place in nonhuman primates [14] and, in this particular case, even in a species that is on a different branch of the evolutionary tree to humans.

In summary, there is substantial evidence to support the assumption that nonhuman primates possess at least rudimentary versions of the fundamental computational mechanisms underlying human language: auditory object recognition and combination in the ventral stream (including a mapping from acoustic input to conceptual representations) and basic sequence processing in the dorsal stream. This suggests that a nonhuman primate (e.g., rhesus macaque) model may indeed constitute a suitable animal model for the neurobiology of higher-order human language processing.

On necessity: Interacting streams as the basis for all aspects of human language

Assuming that our argument for sufficiency as advanced in the previous section can be upheld, the rapidly increasing literature on parallels between birdsong and human language begs another, perhaps even more difficult question: Does the auditory dual-streams architecture of the nonhuman primate model constitute a necessary prerequisite for the emergence of language? Obviously, a definitive answer to this question will remain outstanding unless a species is discovered that has a human-like language system with a distinct neural architecture to support it. Nevertheless, we propose that the dual-streams architecture of the nonhuman primate auditory system does offer some unique advantages that render its necessity for language at least a possibility. Specifically, we assume that the

role of prefrontal cortex in information integration between the dorsal and ventral streams may be of crucial importance in this regard.

We have suggested above that cross-stream interaction is crucial for language and assumed that it is mainly implemented as a transformation of information from one stream to produce a top-down influence on the other, with prefrontal cortex playing a crucial role in this transformation. In this view, the rich expressive power of language is grounded in a neural architecture that accomplishes the integration of auditory object recognition and combination with sequential prediction, and that the prefrontal cortex mediates this ventral/dorsal integration. (For a somewhat similar perspective, though we do not subscribe to the specific linguistic assumptions of this proposal, see [93]).

Thus, perhaps the most promising explanation for the striking differences between humans and nonhuman primates is not within a single stream, but in cross-stream interaction. Anatomical differences in prefrontal cortex between nonhuman primates and humans (e.g., [94]) can be expected to play a crucial role in resolving the question of why humans have language and nonhuman primates do not. While we assume that the relevant computational neurobiological functions of the dorsal and ventral streams are already present in nonhuman primates – at least to some degree – the complexity of interactive computation across the two streams may be what yields the vast communicative power of language. Selected predictions following from this approach and outstanding questions are outlined in Box 3 and Box 4, respectively.

Concluding remarks

The model proposed here is conservative in certain respects and radical in certain others. It is conservative in assuming a primate model as the basis for the neurobiology of language, since the neuroanatomy of nonhuman primates has the closest correspondence to that of humans. It is radical for the very same reason: In spite of the obvious and extensive anatomical homologies between humans and nonhuman primates, neurobiological models of higher-order language have sought other candidate animal models (or no such models) to rely on, with songbird models as the most prominent (e.g. [95]), based on the behavioral and assumed computational differences between humans and nonhuman primates in communication and sequence processing. We have suggested that, to the contrary, the basic computational biological prerequisites for human language, including sentence and discourse processing, are already present in nonhuman primates. Across species, the antero-ventral stream performs a mapping from complex spectro-temporal patterns to time-invariant conceptual representations; it subserves the identification of increasingly more complex auditory objects and hence performs dependency formation as a basic computational mechanism in language. The postero-dorsal stream, by contrast, performs sequence processing in accordance with the constraints posed by an internal model. This ability may be rooted in mechanisms for spatial processing and/or action understanding, which have been generalized to apply to other domains. While the functions of the two streams differ quantitatively across species, we have argued that there is currently no compelling evidence for qualitative leaps in either stream between humans and nonhuman primates. Rather, the fascinating communicative power of human language may be rooted in

the ability to share information dynamically across streams. This proposition reconciles the substantial phylogenetic differences in prefrontal cortex, which go hand in hand with the ability for language, with the neurobiologically plausible assumption that prefrontal cortex serves to regulate brain activity and behavior rather than performing domain-specific computational functions.

Acknowledgements

We would like to thank Sarah Tune and Phillip Alday for helpful discussions on this line of research as well as three anonymous reviewers for their constructive feedback on previous versions of this manuscript. The research reported here was supported by grants from the LOEWE program of the German state of Hesse (grant III L 4–518/70.004 to IBS), the German Research Foundation (grants TRR 135/1 C05 to IBS and SCHL544/6-1 to MS), the National Science Foundation (grants BCS-0519127 and OISE-0730255 to JPR) and the National Institutes of Health (grants R01DC03489 and R01NS052494 to JPR; R01DC-R01-3378 and P01HD040605-8228 to SLS) and with partial support of the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement n° 291763 (JPR).

References

1. Hauser MD, et al. The faculty of language: what it is, who has it, and how did it evolve? *Science*. 2002; 298:1569–1579. [PubMed: 12446899]
2. Berwick RC, et al. Evolution, brain, and the nature of language. *Trends Cogn. Sci.* 2013; 17:89–98. [PubMed: 23313359]
3. Rauschecker JP. Cortical processing of complex sounds. *Curr. Op. Neurobiol.* 1998; 8:516–521. [PubMed: 9751652]
4. Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 2009; 12:718–724. [PubMed: 19471271]
5. Hickok G, Poeppel D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 2007; 8:393–402. [PubMed: 17431404]
6. Ueno T, et al. Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*. 2011; 72:385–396. [PubMed: 22017995]
7. Saur D, et al. Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. USA*. 2008; 105:18035–18040. [PubMed: 19004769]
8. Friederici AD. The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.* 2012; 16:262–268. [PubMed: 22516238]
9. Friederici AD. Language development and the ontogeny of the dorsal pathway. *Front. Evol. Neurosci.* 2012; 4:3. [PubMed: 22347185]
10. Bolhuis JJ, et al. Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* 2010; 11:747–759. [PubMed: 20959859]
11. Margoliash D, Nusbaum HC. Language: the perspective from organismal biology. *Trends Cogn. Sci.* 2009; 13:505–510. [PubMed: 19892586]
12. Petkov C, Jarvis ED. Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* 2012; 4:12. [PubMed: 22912615]
13. Fitch WT. The evolution of syntax: an exaptationist perspective. *Front. Evol. Neurosci.* 2011; 3:9. [PubMed: 22207847]
14. Ghazanfar AA, Takahashi DY. The evolution of speech: vision, rhythm, cooperation. *Trends Cogn. Sci.* 2014; 18:543–553. [PubMed: 25048821]
15. Van Valin, RD, Jr. Exploring the Syntax-Semantics Interface. Cambridge University Press; 2005.
16. Bresnan, J. Lexical Functional Grammar. Blackwell; 2001.
17. Chomsky, N. Minimalist Inquiries: The Framework. In: Martin, R.; Michaels, D.; Uriagereka, J., editors. Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik. MIT Press; 2000. p. 89-155.

18. Goldberg AE. Constructions: a new theoretical approach to language. *Trends Cogn. Sci.* 2003; 7:219–224. [PubMed: 12757824]
19. Rauschecker JP. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Res.* 2011; 271:16–25.
20. Rauschecker JP, Tian B. Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. USA.* 2000; 97:11800–11806. [PubMed: 11050212]
21. Bizley JK, Cohen YE. The what, where and how of auditory-object perception. *Nat Rev Neurosci.* 2013; 14:693–707. [PubMed: 24052177]
22. DeWitt I, Rauschecker JP. Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. USA.* 2012; 109:E505–E514. [PubMed: 22308358]
23. Hasson U, et al. Abstract coding of audiovisual speech: beyond sensory representation. *Neuron.* 2007; 56:1116–1126. [PubMed: 18093531]
24. Mesulam M-M, et al. The core and halo of primary progressive aphasia and semantic dementia. *Ann. Neurol.* 2003; 54:S11–S14. [PubMed: 12833362]
25. Patterson K, et al. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 2007; 8:976–988. [PubMed: 18026167]
26. Gorno-Tempini ML, et al. Classification of primary progressive aphasia and its variants. *Neurology.* 2011; 76:1006–1014. [PubMed: 21325651]
27. Martin A. The representation of object concepts in the brain. *Annu. Rev. Psychol.* 2007; 58:25–45. [PubMed: 16968210]
28. Desai RH, et al. The neural career of sensory-motor metaphors. *J. Cogn. Neurosci.* 2011; 23:2376–2386. [PubMed: 21126156]
29. Visser M, et al. Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J. Cogn. Neurosci.* 2010; 22:1083–1094. [PubMed: 19583477]
30. Pobric G, et al. Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proc. Natl. Acad. Sci. USA.* 2007; 104:20137–20141. [PubMed: 18056637]
31. Chiou R, et al. A conceptual lemon: Theta burst stimulation to the left anterior temporal lobe untangles object representation and its canonical color. *J. Cogn. Neurosci.* 2014; 26:1066–1074. [PubMed: 24345174]
32. Brennan J, et al. Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain Lang.* 2012; 120:163–173. [PubMed: 20472279]
33. Bemis DK, Pylkkanen L. Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cereb. Cortex.* 2013; 23:1859–1873. [PubMed: 22735156]
34. Bornkessel-Schlesewsky I, Schlesewsky M. Reconciling time, space and function: A new dorsal-ventral stream model of sentence comprehension. *Brain Lang.* 2013; 125:60–76. [PubMed: 23454075]
35. Mishkin M, et al. Object vision and spatial vision: Two cortical pathways. *Trends Neurosci.* 1983; 6:414–417.
36. Goodale MA, Milner D. Separate visual pathways for perception and action. *Trends Neurosci.* 1992; 15:20–25. [PubMed: 1374953]
37. Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 2000; 4:131–138. [PubMed: 10740277]
38. Hickok G, et al. Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron.* 2011; 69:407–422. [PubMed: 21315253]
39. Wise RJS, et al. Separate neural subsystems within 'Wernicke's area'. *Brain.* 2001; 124:83–95. [PubMed: 11133789]
40. Scott SK, Wise RJS. The functional neuroanatomy of prelexical processing in speech perception. *Cognition.* 2004; 92:13–45. [PubMed: 15037125]
41. Benson RR, et al. Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain Lang.* 2001; 78:364–396. [PubMed: 11703063]

42. Geiser E, et al. The neural correlate of speech rhythm as evidenced by metrical speech processing. *J. Cogn. Neurosci.* 2008; 20:541–552. [PubMed: 18004944]
43. Skipper, JI., et al. Lending a helping hand to hearing: another motor theory of speech perception. In: Arbib, MA., editor. *Action to Language via the Mirror Neuron System*. Cambridge University Press; 2006. p. 250-285.
44. Jordan MI, Rumelhart DE. Forward models: Supervised learning with a distal teacher. *Cogn. Sci.* 1992; 16:307–354.
45. Wolpert DM, Ghahramani Z. Computational principles of movement neuroscience. *Nat. Neurosci.* 2000; 3:1212–1217. [PubMed: 11127840]
46. Friston KJ. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* 2010; 11:127–138. [PubMed: 20068583]
47. Pickering MJ, Garrod S. An integrated theory of language production and comprehension. *Behav. Brain Sci.* 2013; 36:329–347. [PubMed: 23789620]
48. Pickering MJ, Clark A. Getting ahead: forward models and their place in cognitive architecture. *Trends Cogn. Sci.* 2014; 18:451–456. [PubMed: 24909775]
49. Fiebach CJ, et al. Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. *Hum. Brain Mapp.* 2005; 24:79–91. [PubMed: 15455462]
50. Bornkessel I, et al. Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage.* 2005; 26:221–233. [PubMed: 15862222]
51. Bornkessel-Schlesewsky I, et al. Word order and Broca's region: Evidence for a supra-syntactic perspective. *Brain Lang.* 2009; 111:125–139. [PubMed: 19853290]
52. Friston K. A theory of cortical responses. *Phil. Trans. R. Soc. B.* 2005; 360:815–836. [PubMed: 15937014]
53. Wolpert DM, et al. A unifying computational framework for motor control and social interaction. *Phil Trans. R. Soc. B.* 2003; 358:593–602. [PubMed: 12689384]
54. Haruno M, et al. Hierarchical MOSAIC for movement generation. *Intl. Congress Series.* 2003; 1250:575–590.
55. Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 1999; 2:79–87. [PubMed: 10195184]
56. Pallier C, et al. Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. USA.* 2011; 108:2522–2527. [PubMed: 21224415]
57. Bever, TG. The cognitive basis for linguistic structures. In: Hayes, JR., editor. *Cognition and language development*. Wiley; 1970. p. 279-362.
58. Wilson SM, et al. Syntactic processing depends on dorsal language tracts. *Neuron.* 2011; 72:397–403. [PubMed: 22017996]
59. Wilson SM, et al. What Role Does the Anterior Temporal Lobe Play in Sentence-level Processing? Neural Correlates of Syntactic Processing in Semantic Variant Primary Progressive Aphasia. *J. Cogn. Neurosci.* 2014; 26:970–985. [PubMed: 24345172]
60. Miller E, Cohen JD. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 2001; 24:167–202. [PubMed: 11283309]
61. Cloutman LL. Interaction between dorsal and ventral processing streams: where, when and how? *Brain Lang.* 2013; 127:251–263. [PubMed: 22968092]
62. Zanon M, et al. Cortical connections between dorsal and ventral visual streams in humans: Evidence by TMS/EEG co-registration. *Brain Topogr.* 2010; 22:307–317. [PubMed: 19499197]
63. Lieberman P. The evolution of human speech. *Curr. Anthropol.* 2007; 48:39–66.
64. Hagoort P. MUC (Memory, Unification, Control) and beyond. *Front. Psychol.* 2013; 4:416. [PubMed: 23874313]
65. Thompson-Schill SL, et al. The frontal lobes and the regulation of mental activity. *Curr. Op. Neurobiol.* 2005; 15:219–224. [PubMed: 15831406]
66. Koechlin E, Summerfield C. An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.* 2007; 11:229–235. [PubMed: 17475536]
67. Badre D, D'Esposito M. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat. Rev. Neurosci.* 2009; 10:659–669. [PubMed: 19672274]

68. Terrace HS, et al. Can an ape create a sentence? *Science*. 1979; 206:891–902. [PubMed: 504995]
69. Fitch WT. The evolution of speech: a comparative review. *Trends Cogn. Sci.* 2000; 4:258–267. [PubMed: 10859570]
70. Suga N, et al. Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the mustache bat. *Science*. 1978; 200:778–781. [PubMed: 644320]
71. Margoliash D, Fortune ES. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. *J. Neurosci.* 1992; 12:4309–4326. [PubMed: 1432096]
72. Rauschecker JP, et al. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*. 1995; 268:111–114. [PubMed: 7701330]
73. Friederici AD, et al. The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. USA.* 2006; 103:2458–2463. [PubMed: 16461904]
74. Whiten A. Imitation of the sequential structure of actions by chimpanzees (Pan troglodytes). *J. Comp. Psychol.* 1998; 112:270–281. [PubMed: 9770315]
75. Chalmers M, McGonigle B. Are children any more logical than monkeys on the five-term series problem? *J. Exp. Child Psychol.* 1984; 37:355–377.
76. D'Amato MR, Colombo M. The symbolic distance effect in monkeys (*Cebus apella*). *Anim. Learn. Behav.* 1990; 18:133–140.
77. Feigenson L, et al. Core systems of number. *Trends. Cogn. Sci.* 2004; 8:307–314. [PubMed: 15242690]
78. Jensen G, et al. Transfer of a serial representation between two distinct tasks by Rhesus macaques. *PLoS ONE.* 2013; 8:e70285. [PubMed: 23936179]
79. Lewis JW, Van Essen DC. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 2000; 428:112–137. [PubMed: 11058227]
80. Petrides M, Pandya DN. Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 2009; 7:e1000170. [PubMed: 19668354]
81. Romanski LM, et al. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neurosci.* 1999; 2:1131–1136. [PubMed: 10570492]
82. Frey S, et al. Cortico-cortical connections of areas 44 and 45B in the macaque monkey. *Brain Lang.* 2014; 131:36–55. [PubMed: 24182840]
83. Kohler E, et al. Hearing sounds, understanding actions: action representation in mirror neurons. *Science*. 2002; 297:846–848. [PubMed: 12161656]
84. Artchakov D, et al. Representation of sound sequences in the auditory dorsal stream after sensorimotor learning in the rhesus monkey. *Soc. Neurosci. Abstr.* 2012 368.04.
85. Kaas JH. The organization of neocortex in mammals: Implications for theories of brain function. *Annu. Rev. Psychol.* 1987; 38:129–151. [PubMed: 3548573]
86. Van Essen DC, Dierker DL. Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron*. 2007; 56:209–225. [PubMed: 17964241]
87. Conway CM, Christiansen MH. Sequential learning in non-human primates. *Trends Cogn. Sci.* 2001; 5:539–546. [PubMed: 11728912]
88. MacNeilage PF. The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 1998; 21:499–511. [PubMed: 10097020]
89. Ghazanfar AA, et al. Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Curr. Biol.* 2012; 22:1176–1182. [PubMed: 22658603]
90. Chandrasekaran C, et al. The natural statistics of audiovisual speech. *PLoS Comp. Biol.* 2009; 5:e1000436.
91. Ghazanfar AA, Takahashi DY. Facial Expressions and the Evolution of the Speech Rhythm. *J Cogn. Neurosci.* 2014; 26:1196–1207. [PubMed: 24456390]
92. Takahashi DY, et al. Couple oscillator dynamics of vocal turn-taking in monkeys. *Curr. Biol.* 2013; 23:2162–2168. [PubMed: 24139740]
93. Miyagawa S, et al. The Integration Hypothesis of Human Language Evolution and the Nature of Contemporary Languages. *Front. Psychol.* 2014; 5:564. [PubMed: 24936195]

94. Passingham, RE.; Wise, SP. The neurobiology of the prefrontal cortex: Anatomy, evolution, and the origin of insight. Oxford University Press; 2012.
95. Doupe AJ, Kuhl PK. Birdsong and human speech: Common themes and mechanisms. *Annu. Rev. Neurosci.* 1999; 22:567–631. [PubMed: 10202549]
96. Tian B, et al. Functional specialization in rhesus monkey auditory cortex. *Science.* 2001; 292:290–293. [PubMed: 11303104]
97. Pettito, L-A. How the brain begets language. In: McGilvray, J., editor. *The Cambridge Companion to Chomsky.* Cambridge University Press; 2005. p. 84-101.
98. Bisang W. Precategoriality and syntax-based parts of speech: The case of Late Archaic Chinese. *Studies Lang.* 2008; 32:568–589.
99. Gil D. Creoles, complexity and Riau Indonesian. *Ling. Typol.* 2001; 5:325–371.
100. Christiansen MH, Chater N. Towards a connectionist model of recursion in linguistic performance. *Cogn. Sci.* 1999; 23:157–205.
101. Collier K, et al. Language evolution: syntax before phonology? *Proc. R. Soc. B.* 2014; 281:20140263.
102. Everett DL. Cultural constraints on grammar and cognition in Pirahã. *Curr. Anthropol.* 2005; 46:621–646.
103. Nevins A, et al. Pirahã exceptionality: A reassessment. *Language.* 2009; 85:355–404.
104. MacWhinney B, et al. Cue validity and sentence interpretation in English, German and Italian. *J. Verb. Learn. Verb. Behav.* 1984; 23:127–150.
105. Bornkessel-Schlesewsky I, et al. Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain Lang.* 2011; 117:133–152. [PubMed: 20970843]
106. Tune S, et al. Cross-linguistic variation in the neurophysiological response to semantic processing: Evidence from anomalies at the borderline of awareness. *Neuropsychologia.* 2014; 56:147–166. [PubMed: 24447768]
107. Dehaene S, Cohen L. Cultural recycling of cortical maps. *Neuron.* 2007; 56:384–398. [PubMed: 17964253]
108. Dehaene S, Cohen L. The unique role of the visual word form area in reading. *Trends. Cogn. Sci.* 2011; 15:254–262. [PubMed: 21592844]
109. Ghazanfar AA, et al. Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: implications for the evolution of audiovisual speech. *E. J. Neurosci.* 2010; 31:1807–1817.
110. Ackermann H, et al. Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *Behav. Brain Sci.* 2014:1–84.
111. Leaver AM, et al. Brain activation during anticipation of sound sequences. *J. Neurosci.* 2009; 29:2477–2485. [PubMed: 19244522]
112. Pasupathy A, Miller EK. Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature.* 2005; 433:873–876. [PubMed: 15729344]
113. Petkov C, et al. A voice region in the monkey brain. *Nat. Neurosci.* 2008; 11:367–374. [PubMed: 18264095]
114. Lerner Y, et al. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* 2011; 31:2906–2915. [PubMed: 21414912]

Box 1**Neurobiological design principles and the computational division of labor between the antero-ventral and postero-dorsal streams**

The assumption that the auditory cortical system of (nonhuman) primates (with its extensions into parietal, premotor and prefrontal cortices) provides a suitable animal model for the neurobiology of human language allows us to adopt several neurobiological design principles that have been tested empirically within the animal (primarily rhesus macaque) model. The two most important of these are:

- **Hierarchical organization of the processing streams.** Hierarchical organization implies increasing combination sensitivity to individual stimulus features with increasing distance from primary sensory cortices (here: from auditory cortex, AC). Single-unit studies in rhesus monkeys have demonstrated hierarchical organization of the ventral auditory stream [72,96], with neurons in primary auditory cortex A1 and in the middle lateral belt region of AC processing elementary auditory features (e.g., frequency-modulated (FM) sweeps or bandpass noise bursts) and neurons in more anterior portions of the lateral belt responding to more complex spectro-temporal patterns (e.g. species-specific vocalizations, monkey calls).
- **Separable but internally unified computational properties for each stream.** The traditional subdivision of dual (visual or auditory) streams into a "what" and a "where" (or "how") stream suggests a computational division of labor between streams with a common computational denominator within each stream [20,35,36]. We adopt this assumption and posit that the major dorsal-ventral distinction is functionally relevant irrespective of the possible presence of anatomical sub-pathways (e.g. multiple white-matter tracts within one stream).

The functional division of labor between the antero-ventral and postero-dorsal streams envisaged here is illustrated in Figure I. The antero-ventral stream (Figure Ia) recognizes successively more complex auditory objects. As described in the main text for linguistic dependency formation (i.e. the establishment of order-independent interpretive relations between elements), feature combination in more complex auditory objects is assumed to be commutative (i.e., order insensitive). Order insensitivity is a crucial property of conceptual representations, since these can be identified via varying sensory stimuli (auditory, visual, olfactory, etc.), and the combination of individual features to form a concept is not dependent on a particular order of features in the input (e.g. has-four-legs, has-a-tail, barks as features of the concept "dog" are not ordered with respect to one another). The postero-dorsal stream (Figure Ib), by contrast, performs predictive sequence processing via hierarchically organized internal models. Prediction at one hierarchical level is assumed to influence predictive processing at the next-lower level via feedback connections, eventually leading to a comparison of predicted sensory information with the actual sensory input. Error signals resulting from this comparison are transmitted upwards in the hierarchy via feedforward connections [46,52].

Box 2**Proposed computational differences between humans and nonhuman primates reconsidered**

Behavioral differences between humans and nonhuman primates have played a major role in shaping the debate regarding a suitable animal model for the neurobiology of language. Upon closer consideration, however, some of the key differences that have been posited in this regard may not be as substantial as typically assumed.

Ventral stream

With regard to auditory object perception and production, language-learning experiments with chimpanzees have led some researchers to conclude that “chimps do not really have ‘names for things’ at all” on the basis of the fact that they “will use the same label apple to refer to the action of eating apples, the location where apples are kept [...]” [97]. This property (“transcategoriality”), however, is also true of a number of human languages, including Classical Chinese [98], Riau Indonesian [99] and other Austronesian languages. Possible auditory objects and their functions must thus be considered in the face of the full range of human languages (~ 7000).

Dorsal stream

Neuroscientists interested in the neurobiology of complex language structures often base their line of enquiry on one specific, assumed aspect of language competence: the capacity for discrete infinity [1]. We argue that this notion is based on a logical calculus that abstractly characterizes human grammars, rather than on empirically substantiated, biologically-motivated hypotheses (see also [11]). Humans themselves are actually quite severely limited in the number of center embeddings that they can process, cf. the difficulty arising in “If if the cat is in, then the dog cannot come in, then the cat and dog dislike each other” [100]. Note that this sentence is analogous to the AABB structure that is commonly used in experiments on the neural correlates of recursive (“human”) grammars (e.g. [73]). Once actual human performance is considered the benchmark, these can be processed by considerably simpler computational mechanisms than those posited as necessary in the majority of neuroscientific examinations of human vs. nonhuman sequence processing abilities [100]. Hence, if human abilities are measured in the same way as those of nonhuman primates or birds necessarily must be, i.e. by means of behavioral performance, then cross-species differences can be described almost entirely in quantitative rather than qualitative terms. For converging comparative evidence and the conclusion that syntactic structure as present in human language may in fact be more widespread throughout animal communication systems than phonological structure, see [101].

Box 3**Selected predictions arising from the model put forward here**

- **Languages with only a limited (or even no) capacity for recursion.** The assumption that computational differences between audition in nonhuman primates and language in humans are quantitative rather than qualitative leads to the prediction that, though recursion is a property of most human languages, a small number of languages at the outer edge of the distribution may show only a limited capacity for recursion. This prediction indeed appears to be borne out, e.g. in the much discussed case of the Pirahã language spoken in the Amazon region of Brazil ([102]; but see e.g. [103]).
- **Neurobiological differences between sequence-based and non-sequence-based language processing.** The computational functions of the dorsal and ventral streams posited here predict that the default division of labor between streams may differ depending on language-specific characteristics. For example, as already noted in the main text, it is well established that some languages are more sequence-based than others (e.g. [104]). Accordingly, the model put forward here predicts that strongly sequence-based (e.g. English) versus weakly sequence-based languages (e.g. German, Turkish) may rely on a different organization of the underlying neurobiological processing architecture. Initial evidence from the electrophysiological domain suggests that this prediction is borne out [105,106].
- **Extension to reading.** We claim here that the neurobiological basis of language is rooted in primate higher audition and the computational division of labor between the two auditory processing streams. By contrast, there is no evidence for a biological basis of reading or writing in any other species. This leads to the following two predictions. Firstly, reading will likely rely on neural mechanisms (dorsal and ventral) that largely overlap with (or extend from) those described here for (auditory) language processing. (Note that all current neurobiological models of reading focus on the word level. Thus, perhaps surprisingly, this claim is largely untested for higher levels, such as sentence reading). Secondly, the differing input system obviously drawn upon during reading must necessarily have been exapted from another neural system (presumably within the visual ventral stream). This second prediction is compatible with the neuronal recycling hypothesis [107,108], which postulates the re-purposing of existing brain maps for culturally recent tasks.

Box 4**Outstanding questions**

- To what extent can the model proposed here be generalized to language production? We have emphasized comprehension in the present paper, partly in view of the obvious discrepancies between the vocal abilities of humans and nonhuman primates and the resulting insight that computational parallels may lie elsewhere. However, the perspective that there may be a close link between human speech processing and the perception and production of rhythmic orofacial movements in nonhuman primates [88,89,91,109] also opens up potentially new avenues of investigation for connections at the production level.
- What role do subcortical structures and cortico-subcortical loops play in the dual-streams architecture advocated here? It has been suggested that the phylogenetic development of cortico-striatal circuits may have played a crucial role in the evolution of spoken language [110]. From the perspective of the current model, this proposal provides a possible explanation for the gap between nonhuman primate vocal abilities and the claim that they possess the necessary computational building blocks for language processing. According to an alternative theory, however, subcortical structures - and particularly the basal ganglia - are not only crucial for speech, but also contribute critically to basic sequencing processes [63], and indeed, the basal ganglia have been shown to play an important role in sequence learning both in humans [111] and monkeys [112].
- How does the model proposed here relate to differences between different nonhuman primate species? In this paper, the nonhuman primate model under discussion has primarily been a rhesus macaque model, based on the fact that most of the critical neurobiological nonhuman primate data on which our model is based were obtained in studies with macaques. However, a more detailed future proposal would clearly need to take into account any differences between different nonhuman primate species, e.g. with respect to the size and cytoarchitectonic structure of prefrontal cortex, which we have emphasized here as being crucial for cross-stream integration. Such data should increasingly become available with the use of noninvasive neuroimaging techniques in nonhuman primates (see [12,113]).

Highlights

- We present a unified dual-streams model for primate audition and human language
- Dorsal/ventral stream computations in language are grounded in primate audition
- Ventral stream: auditory object recognition and commutative combinatorics
- Dorsal stream: hierarchical predictive coding of sequences from phonemes to discourse

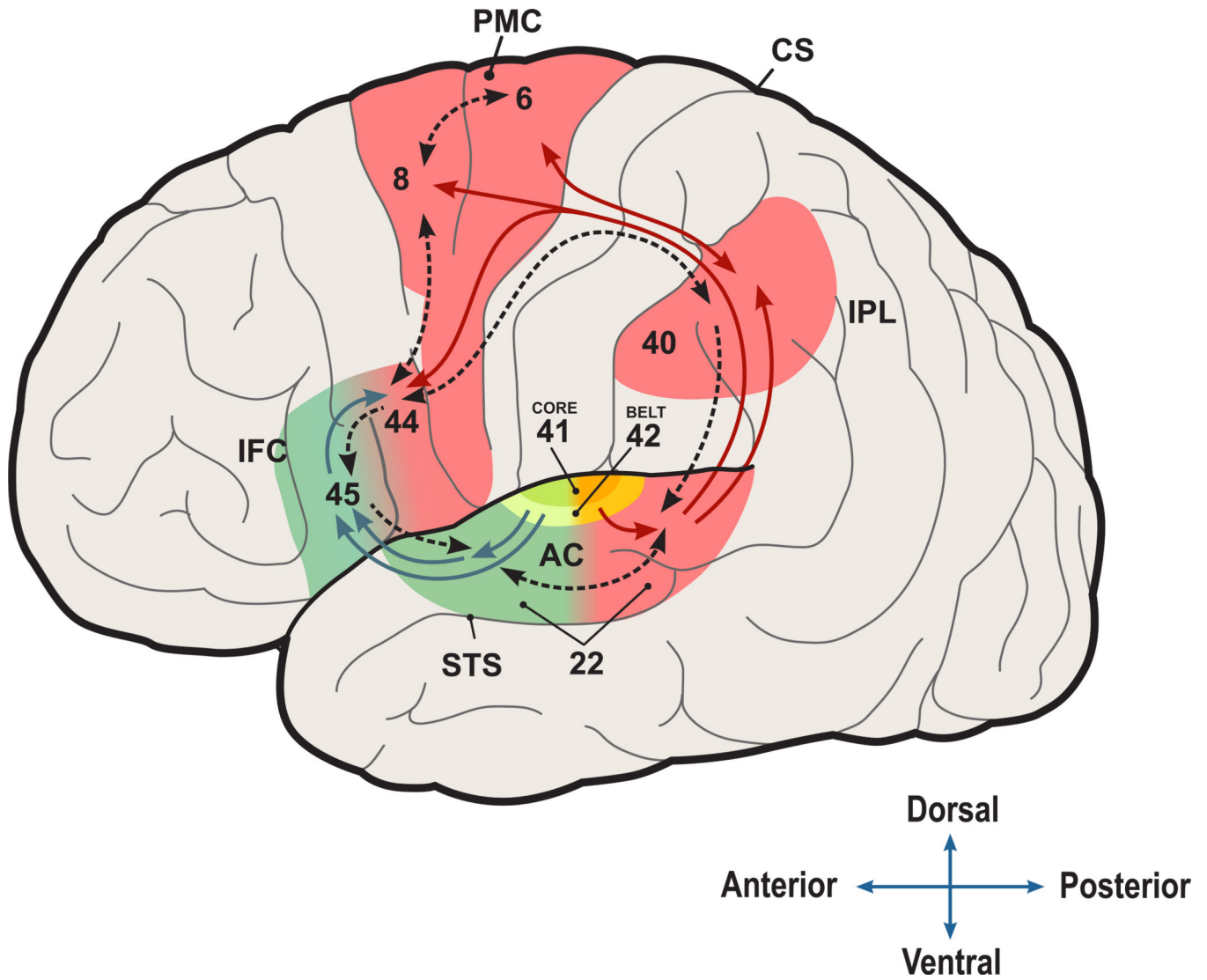
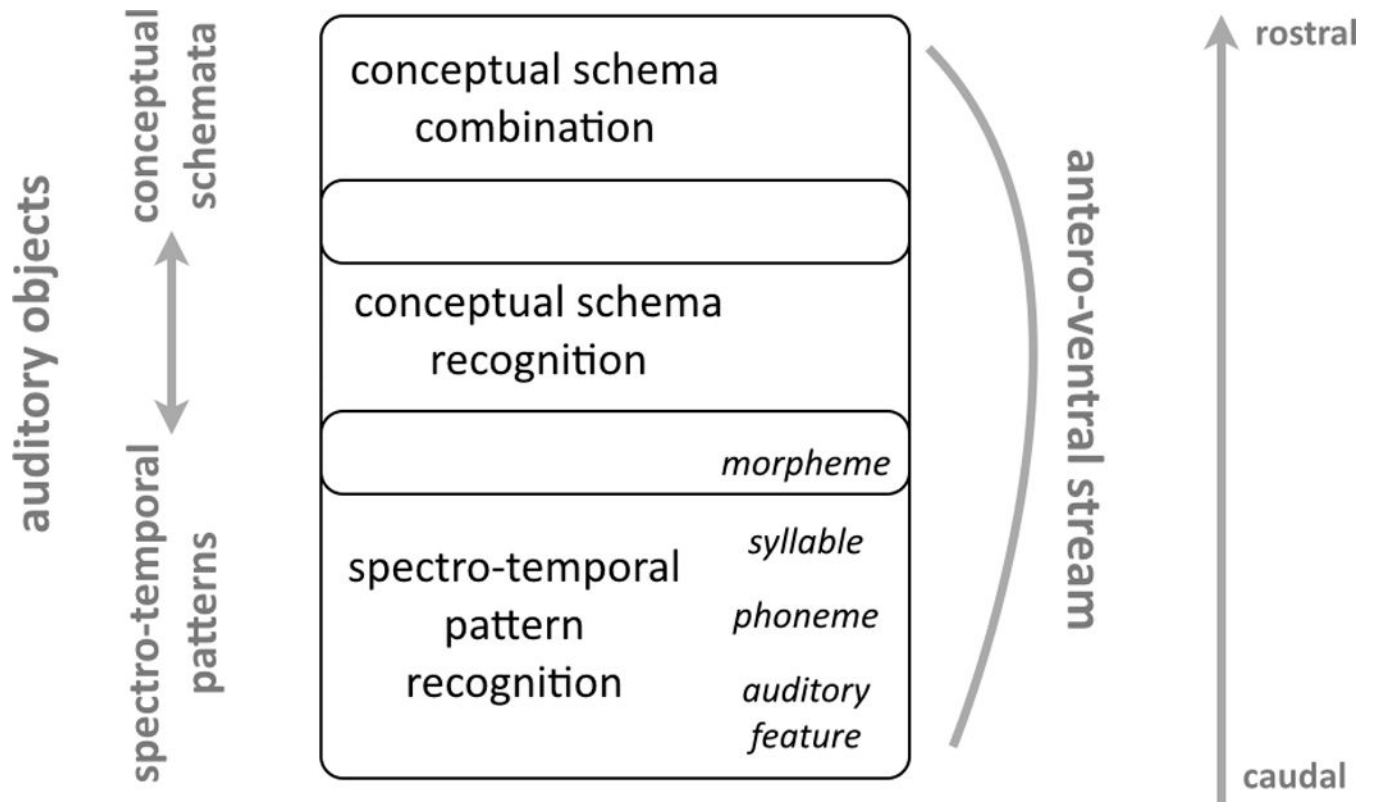


Figure 1. Dual streams supporting language processing in the human brain

Multiple streams of information processing underlie the human ability for speech and language. Neuroanatomically, two streams, a (postero-) dorsal (red) and an (antero-)ventral (green) stream are particularly important, and show clear homologies to the nonhuman primate brain and specifically to its auditory cortical system with its extensions into parietal, premotor and prefrontal cortex. The dorsal stream, in particular, serves to connect sensory (auditory) and motor-related regions of the cerebral cortex and thus implements bidirectional perception-action loops. Cross-stream integration in the inferior frontal gyrus is depicted by the transition from red to green. Black dashed arrows denote feedback connections. Figure adapted from [4].

Abbreviations: AC, auditory cortex; CS, central sulcus; IFC, inferior frontal cortex; IPL, inferior parietal lobule; PMC, premotor cortex. Numbers denote Brodmann areas.



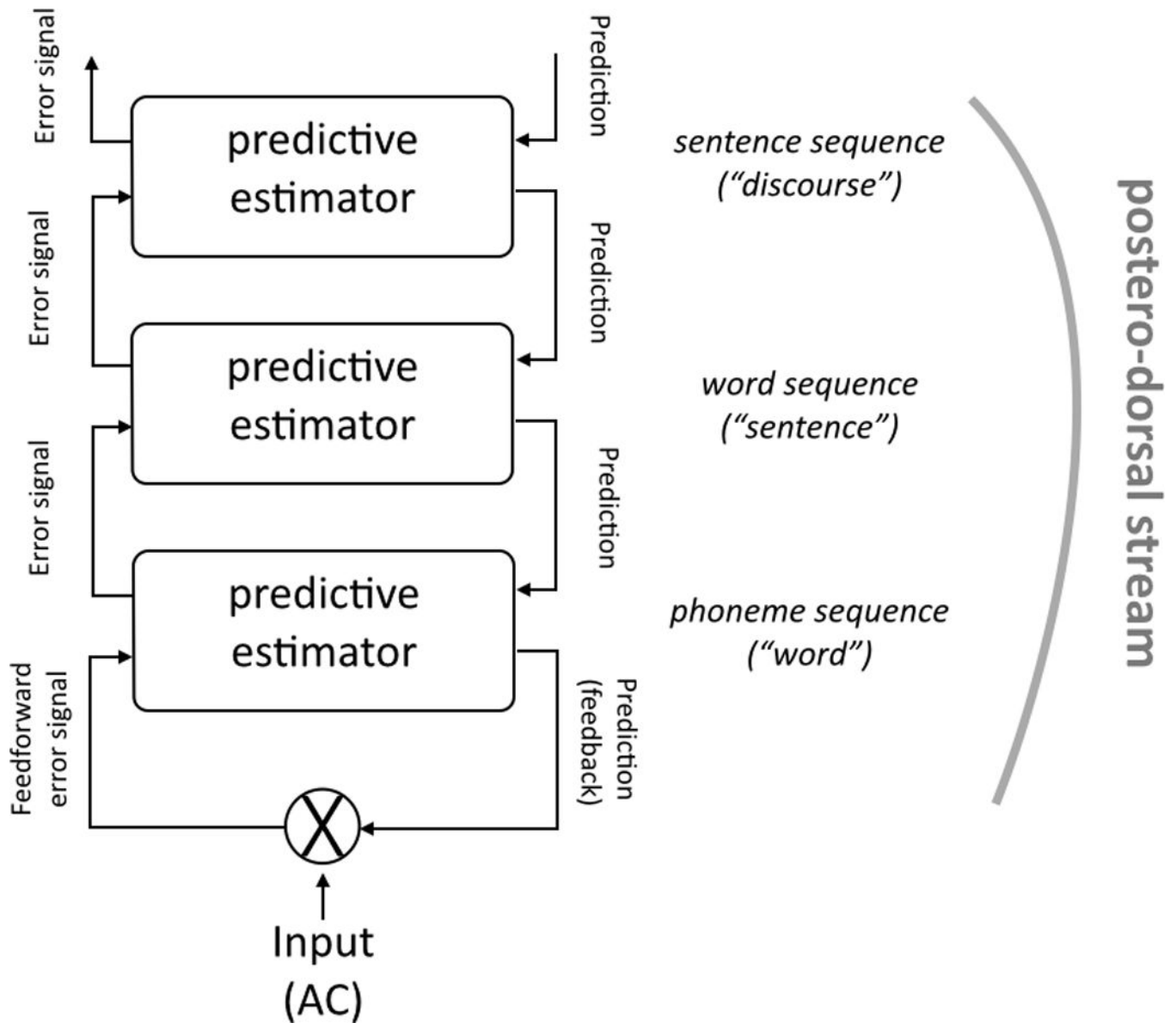


Figure I.

A: Computational properties of the antero-ventral stream Simplified depiction of the computational properties of the antero-ventral stream, which detects auditory objects in a hierarchically organized manner. Thus, neuronal populations close to auditory cortex (AC) detect elementary auditory features, while neuronal populations further downstream from AC detect auditory objects defined via more complex feature combinations. This stream therefore performs a transformation from auditory feature detection to conceptual feature detection (i.e., conceptual schemata are viewed as more complex and abstract auditory objects are viewed as bundles of conceptual features). We hypothesize that, in language, the transition from auditory to conceptual processing may take place at the level of the morpheme as the smallest meaning-bearing unit. "Auditory" object detection is thus predicted to be independent of sensory modality from the morpheme level upwards.

B. Computational properties of the postero-dorsal stream Simplified depiction of the computational properties of the postero-dorsal stream, which performs sequence processing in accordance with a hierarchically organized set of internal models. In contrast to the antero-ventral stream, feature combinations forming successively more complex (i.e., longer and hierarchically deeper) sequences are order-sensitive (non-commutative). Predictions generated at one hierarchical level provide a feedback (top-down) influence on the next level down in the hierarchy, while error signals arising from the comparison between the predicted and actual sensory input are transmitted up the hierarchy via feedforward connections. The three hierarchical levels (word, sentence, discourse) shown here are based on the empirical finding of successively larger auditory integration windows comprising these levels in the dorsal auditory stream [114].
Figure adapted with permission from [54].