

## REVIEW

# Evolution of vocal learning and spoken language

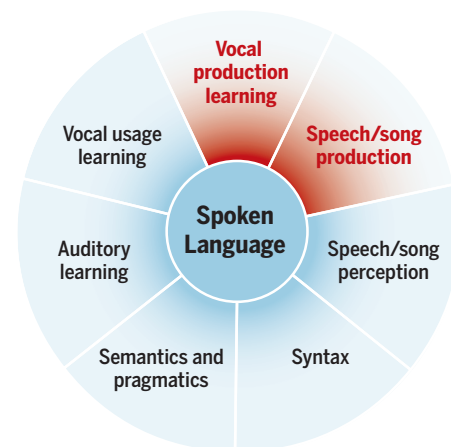
Erich D. Jarvis<sup>1,2</sup>

Although language, and therefore spoken language or speech, is often considered unique to humans, the past several decades have seen a surge in nonhuman animal studies that inform us about human spoken language. Here, I present a modern, evolution-based synthesis of these studies, from behavioral to molecular levels of analyses. Among the key concepts drawn are that components of spoken language are continuous between species, and that the vocal learning component is the most specialized and rarest and evolved by brain pathway duplication from an ancient motor learning pathway. These concepts have important implications for understanding brain mechanisms and disorders of spoken language.

The faculty of spoken language can be thought of as consisting of multiple component traits, some ubiquitous among species and a few specialized to rare groups of species or just to humans (Fig. 1) (1). A ubiquitous component is auditory learning, the ability to learn and remember novel sound associations (2); for example, a dog or nonhuman primate can learn the meaning of the sound “sit” or of word combinations such as “come here, boy” or “play the flute” (3). Also ubiquitous is vocal usage learning, the ability to learn to produce innate or learned sounds in unfamiliar contexts (2)—for example, a dog’s or a nonhuman primate’s ability to learn to bark or call when requesting food or signaling alarm. Rarer is vocal production learning, or simply vocal learning, the ability to imitate sounds (2), which has thus far been found only in humans, cetaceans, pinnipeds, bats, and elephants among mammals, and songbirds, parrots, and hummingbirds among birds (2). Each of these groups has closely related species that lack vocal learning, suggesting independent evolution of the trait (4). For other components, syntax, rules governing sequences of sound, semantics, and pragmatics are also present in nonhuman animals but are more advanced in humans (1). For example, black-capped chickadee songbirds have songs with simple syntax for mating and learned calls that signal predators and predator size, but they are not known to combine these vocalizations into longer sequences with different meanings (5) or have hierarchical syntax. Here, I present a comparative synthesis of spoken language as a form of learned forebrain sensory-motor communication, with some components found in most vertebrates to varying degrees and a highly specialized advanced vocal learning component found only in a few species, with humans being the most advanced in all components, and with all components combined into one trait. I consider spoken language and speech as equivalent.

## Discrete versus continuum hypotheses of spoken-language components

Language, and therefore spoken language and some of its components, is sometimes presented as all or none: a species either has vocal learning or it does not (4). However, differences can be a matter of degree. For example, rudimentary vocal plasticity and learning have been found in some species assumed to be vocal nonlearners, including mice (6) and nonhuman primates (2, 7). Some great apes have been taught to produce rudimentary sign language: By her 30s, Koko the gorilla was producing more than 1000 American Sign Language (ASL) signs, in word combinations that were not taught to her, and recognizing ~2000 spoken words (8). Although none of these species demonstrate advanced vocal learning, syntax, or semantics, they are not completely lacking. In prior studies,



**Fig. 1. Multicomponent view of spoken language.**

Spoken language, or speech, is viewed here as a combination of seven component traits. These traits overlap with components studied in linguistics: semantics, pragmatics, syntax, phonology, and morphology. Red text indicates components that are rarest among vertebrates. Most components could be continuous among species (gradient in the center), with humans being the most advanced. [Diagram and components modified from a figure produced by Tecumseh Fitch for language broadly, and used with permission]

my co-workers and I proposed a continuum hypothesis, where different species have varying degrees of vocal learning that evolves in stepwise manner (fig. S1) (2, 6). This hypothesis is plausible when considering large differences in vocal learning complexity among well-established vocal learners, and so it should not be surprising that so-called vocal nonlearners vary too.

Such variation can be influenced by anatomical mechanisms of how sounds are produced, such as learned raspberry lip-smacking in chimpanzees (9) or diaphragm-induced coughing in Koko (3). In contrast, advanced vocal learners have voluntary control of not only the oral-facial articulators (lips, tongue, beak, and jaw) but also the larynx (in mammals) or syrinx (in birds) (10, 11). The continuum hypothesis does not negate that advanced vocal learning is convergent. Further, different continuums for different components could explain why auditory learning is more advanced than vocal learning in many species, why it is easier to listen (receptive language) than to speak (productive language) in a second language, and why a nonverbal autistic child has more receptive than productive spoken language.

## What anatomy makes vocal learning and spoken language special?

The search for the biological substrates of spoken language and vocal learning in humans and song-learning birds has led to six popular hypotheses (Fig. 2), described below.

### Brain size or neuron density impacts spoken-language circuits

This hypothesis proposes that a larger brain allows more neurons for speech and vocal learning (Fig. 2A). However, brain size is not correlated with vocal learning: hummingbirds, with their tiny brains, can imitate complex vocalizations, whereas chimpanzees, with their much larger brains, cannot (2). Although humans have the largest brains among primates, it is a scaled-up primate brain in size and neuron density (12). Other vocal learning mammals, cetaceans and elephants, have larger brains but also larger bodies (12). In contrast, two of the three vocal learning bird species, songbirds and parrots, have forebrain neuron densities two times that of vocal nonlearning bird species (fig. S2A). Perhaps the higher density in some vocal learning birds and the scaled-up human brain accommodated the space needed for extra neurons of the vocal learning and spoken-language circuits, without losing older circuits and while maintaining brain-to-body size ratios.

### Vocal organ with greater capacity for vocalization diversity

This hypothesis proposes that loss of air sacs and presence of a permanently descended larynx in humans (13) or additional intrinsic syrinx muscles in songbirds (14) endowed them

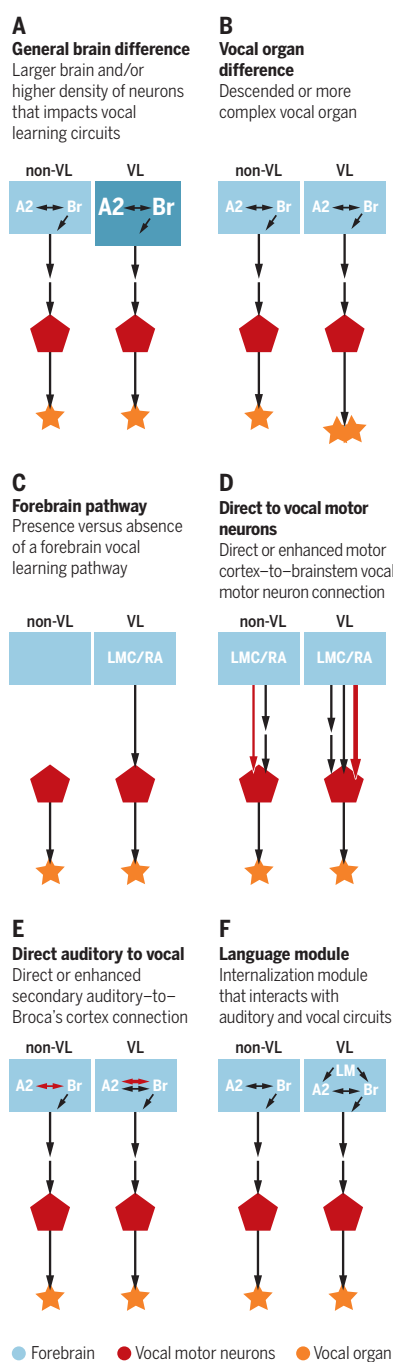
<sup>1</sup>Laboratory of Neurogenetics of Language, The Rockefeller University, New York, NY, USA. <sup>2</sup>Howard Hughes Medical Institute, Chevy Chase, MD, USA.  
Email: ejarvis@rockefeller.edu

with the ability to produce a greater variety of sounds (Fig. 2B). However, subsequent studies found that vocal nonlearning mammals descend their larynx when lifting their heads upward and vocalizing. A nonhuman primate (baboon) larynx can be made in situ to produce the majority of sounds made by the human larynx (fig. S3, A and B). Other nonhuman mammals (lions, koalas, and some ungulates) have independently evolved a permanently descended larynx (13). It is more likely that the air sacs and descended larynx facilitated lower-formant frequencies, allowing an animal to acoustically exaggerate its size (13). Among birds, syrinx muscle complexity does not correlate with vocal learning (fig. S3C), but it is likely that complex vocal musculature allows a vocal nonlearner to produce a greater variety of innate sounds to compensate for lack of forebrain-driven vocal learning (14). Thus, vocal organ differences cannot explain song and speech diversity in vocal learning birds and humans.

#### Forebrain pathways for vocal learning and speech

This hypothesis proposes that only humans and other vocal learning species have a forebrain circuit that controls song and speech (Fig. 2C) (4). All three song-learning bird lineages share seven cerebral nuclei, which make up a posterior vocal pathway for production of learned vocalizations (Fig. 3A, yellow) and an anterior vocal pathway for vocal imitation (Fig. 3A, red) that are not found in vocal nonlearning species (Fig. 3C). I propose that humans have analogous brain regions, which include dorsal and ventral laryngeal motor cortices (dLMC and vLMC) responsible for speech production, and premotor LMC and Broca's area responsible for speech acquisition and higher-level speech functions (Fig. 3B) (4, 10, 15–17). The different songbird song nuclei have cell types that may correspond to the different cortical layers and the striatum and thalamic regions of the human spoken-language pathway (Fig. 4). Input into these specialized song and speech circuits comes from auditory, somatosensory, and other pathways, but these other pathways are found in all vocal nonlearning species investigated to date (Fig. 3) (2, 4), which would explain why auditory learning is more ubiquitous among species. I propose that Wernicke's area and its network involved in speech perception were present in the vertebrate lineage before being elaborated in humans.

In the context of the continuum hypothesis, it has been proposed that nonhuman primates possess a premotor vLMC as the ancestral precursor of human primary vLMC (Fig. 3D) (17, 18). However, premotor vLMC is not required for producing nonhuman primate vocalizations (18), and a rudimentary primary vLMC may already exist in nonhuman primates (Fig. 3D) (19). In mice, a rudimentary LMC was found



**Fig. 2. Hypotheses of anatomically unique features to vocal learning and spoken language.**

Colors indicate brain subdivisions and vocal organs. Arrows indicate neuroanatomical connections; red arrows, continuum hypothesis versions. (A) Larger brain and/or higher density of neurons that affect vocal learning brain circuits. (B) Descended or more complex vocal organ. (C) Presence versus absence of a forebrain vocal learning pathway. (D) Direct or enhanced motor cortex-to-brainstem vocal motor neuron connection. (E) Direct or enhanced secondary auditory-to-Broca's cortex connection. (F) Separate language model. A2, secondary auditory cortex; Br, Broca's area; LM, language module; LMC, laryngeal motor cortex; RA, robust nucleus of the arcopallium; non-VL, nonvocal learner; VL, vocal learner.

with connectivity similar to that of human LMC and the analogous songbird robust nucleus of the arcopallium (RA) (Fig. 3, A and B, and fig. S4). However, unlike those of humans and vocal learning birds, the mouse LMC neurons are embedded in a region that controls non-vocal motor behaviors, does not share specialized gene regulation with vocal learners, is not necessary for producing normal vocalizations, but seems necessary for modulating pitch (6, 15). In birds, a rudimentary RA was found in a suboscine, a close relative of songbirds, but not in quail, a distant relative (20). These findings support the continuum hypothesis of vocal learning. More details on hypothesis 3 are provided in the supplementary text of the supplementary materials.

#### Direct versus indirect motor cortex-to-brainstem vocal motor neuron connection

This hypothesis proposes that a fundamental transition to the evolution of vocal learning and spoken language was a change from (or an addition to) an indirect to a direct projection from human LMC layer 5 neurons and avian RA projection neurons to brainstem vocal motor neurons (Fig. 2D and red arrows in Fig. 3), enabling fine motor control of vocalizations in humans and song-learning birds (1, 2, 18). In the context of the continuum hypothesis, mice LMC layer 5 neurons (6) and the RA-like region in suboscines (20) make sparse direct projections (one to three innervating axons per vocal motor neuron) compared with the dense projections (up to hundreds of innervating axons per vocal motor neuron) in humans and song-learning birds (Fig. 2D and fig. S4). This suggests that density of the direct projection may influence the degree of learned-vocalization production. Experimentally manipulating mice to express less of the repulsive axon guidance receptor *PlexinA1* in layer 5 neurons, to the low levels seen in humans, caused the mice to have a denser direct projection to forelimb motor neurons and greater forelimb manual dexterity (21). Depending on the species, the avian RA song nucleus also makes direct or indirect projections to articulatory (e.g., beak, jaw, tongue) and respiratory motor neurons (fig. S4A). The human oral-facial motor cortex (OMC) between dLMC and vLMC is presumed to do so as well (16). Direct innervation of tongue motor neurons also exists in nonhuman primates (18). This would explain why limited vocal learners have more voluntary control for producing imitated sounds using articulators like the tongue and lips than they do for the larynx.

#### Connection between auditory cortex and speech premotor cortex

This hypothesis proposes that direct connections between secondary auditory cortex (A2 and Wernicke's area) and vocal premotor cortex (preLMC and Broca's area) may endow humans

with the auditory-vocal motor integration necessary to learn, produce, and perceive spoken language (Figs. 2E and 3, blue arrows) (22). A set of dorsal pathways is proposed to control auditory-vocal motor learning for speech (to premotor cortex) and hierarchical syntax for language (to Broca's area), whereas a ventral set is proposed to control lexical and semantic aspects of speech and language (fig. S5A) (22). But different views exist about the connectivity differences between humans and non-human primates (2, 22). Mice were found to contain a direct A2 connection to their LMC-M1/M2 region with respectable density (fig. S5, D to F). Analogous pathways for auditory-vocal motor integration in song-learning birds are those that connect forebrain auditory regions to the forebrain vocal learning nuclei (Fig. 3A, blue arrow) (2). Vocal nonlearning birds are not known to have such forebrain vocal nuclei to project into. It is possible that

different vocal learning lineages evolved different solutions to auditory-vocal motor integration, or that direct connections between auditory and forebrain premotor areas are not specific to humans or vocal learners. Analysis of a greater range of species should bring clarity.

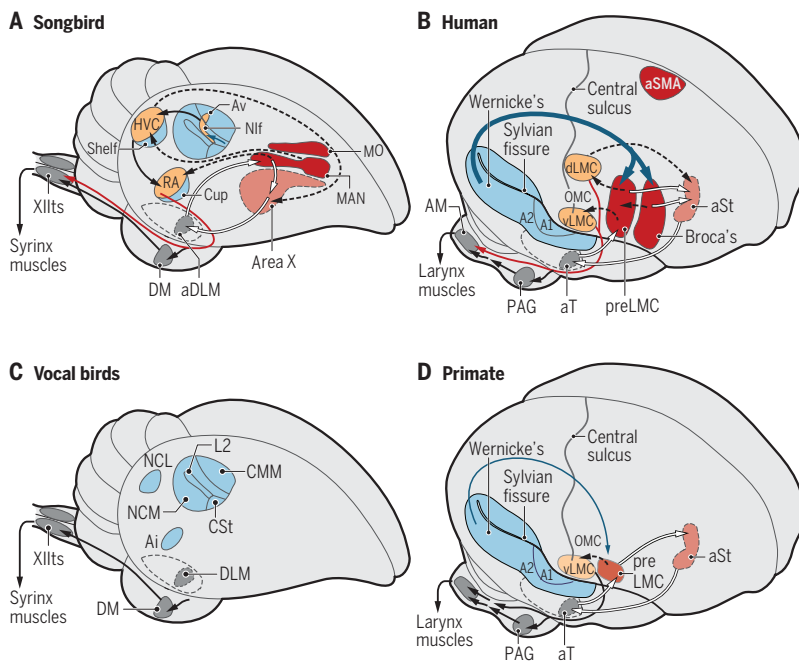
#### Internalization language versus externalization brain circuits

This hypothesis proposes that only humans have an "internalization" language brain circuit, historically thought of as a language module, that processes complex algorithms like hierarchical syntax and the merging of words, which are then expressed through auditory, speech, or limb "externalization" brain circuits shared across species, enabling spoken, signed, and written language in humans (Fig. 2F and fig. S6). A proposed internalization brain region is Broca's area, and a proposed externalization region is speech LMC (22).

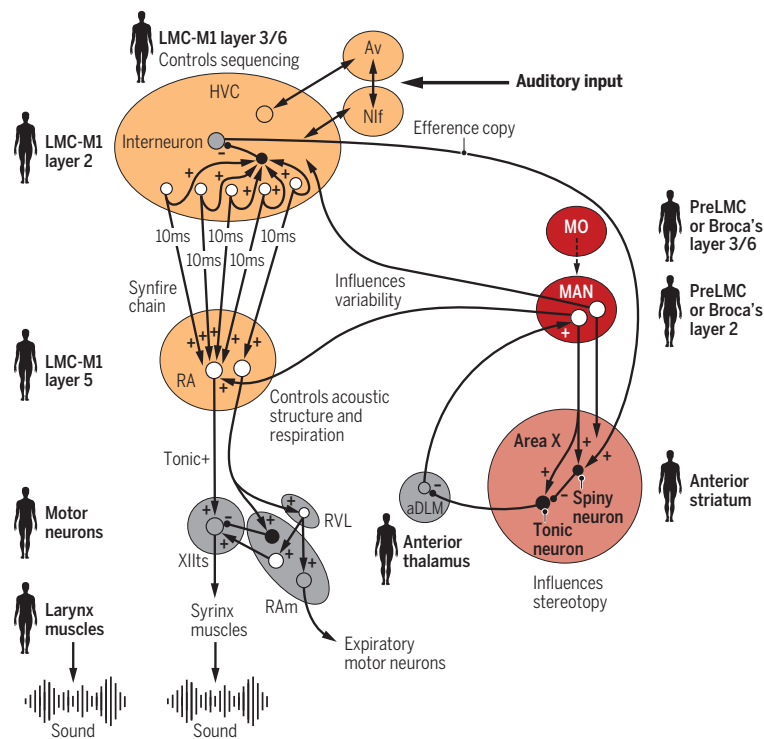
Vocal learning birds are said to only have the externalization circuits, without hierarchical syntax and compositional meaning and without a Broca's analog, yet others have proposed that some of the songbird vocal learning nuclei HVC (letter-based name) or magnocellular nucleus of the anterior nidopallium (MAN) is analogous to Broca's (4). Functional magnetic resonance imaging studies show Broca's is active in language tasks regardless of modality, increasing activity with increasing syntax load (22). In awake patients undergoing surgery, electrical stimulation in Broca's, and also in dLMC and vLMC, can lead to inhibition of ongoing speech and/or hand movements (23); but the same LMC regions (Broca's has not yet been tested) have increased activity mainly during speech-related tasks (10, 16, 24).

It should be noted that most studies like these have not controlled for silent speech production. When we silently speak (i.e., inner speech, thinking in speech), most of the brain areas, including Broca's, used for speech production and perception show increased activity, as do the laryngeal muscles, even though no sound is produced (25). When we hear, read, or write words, activity increases in brain regions for speech production, associated with subvocalization muscle activity of the larynx and other articulators (25). Many limb gestures of ASL are accompanied by "mouth morpheme" movements. As hand-control brain regions are adjacent to vocalization and oral-facial control regions (16, 17), my interpretation of imaging results (24, 26) is that both adjacent regions are active in ASL production as a result of hand and oral movements.

In songbirds, electrophysiology and activity-dependent gene expression studies revealed that the same brain pathways are used to learn and to produce song (Fig. 3A, red and orange) (4). These brain regions and the syrinx muscles show singing neural firing patterns when birds apparently dream about singing (e.g., inner song) (27). There is no "internalization" circuit for song syntax separate from an "externalization" circuit for song production. Adjacent to song learning nuclei are nonvocal motor regions. Translating this to humans: An alternative to the internalization-externalization framework is that the spoken-language brain pathway, inclusive of Broca's, could be used to learn and produce speech, whether voiced or silent during reading, writing, thinking, and signing with mouth morphemes. Nonvocal and nonauditory circuits, e.g., forelimb and vision, may process hierarchical syntax algorithms as the adjacent speech and auditory circuits. This could explain why nonhuman primates have greater sign-language abilities with barely any spoken-language ability. In this view, spoken language and sign language are the same as speech and signing, respectively. These hypotheses can be reconciled with neurophysiological recordings in Broca's and other areas during speech and other tasks.



**Fig. 3. Brain pathways for vocal learning and spoken language.** (A) Vocal learning pathway of songbirds. (B) Vocal learning and spoken-language pathway of humans. (C) Innate brainstem vocal pathway in vocal nonlearning birds. (D) Vocal pathway in nonhuman primates. Comparable brain regions and connections across species are in the same color and projected on smoothed surface brain images. Orange regions and black solid arrows, posterior vocal motor pathway. Red regions and white arrows, anterior vocal pathway. Dashed lines, connections between the two subpathways. Red arrows, specialized direct projection from motor cortex to brainstem vocal motor neurons in vocal learners. Gray regions, innate vocal pathway. Blue regions, auditory regions. Blue arrows, auditory input to specialized vocal learning and spoken-language regions. Subcortical vocal regions are outlined with dashed lines. Orange and red regions in nonhuman primates (D) are less transparent to indicate continuum hypothesis of a rudimentary forebrain vocal circuit. A subset of connections are shown for simplicity. A1, primary auditory cortex; A2, secondary auditory cortex; aDLM, anterior dorsolateral medial nucleus of the thalamus; Ai, intermediate arcopallium; Am, nucleus ambiguus; aSMA, anterior supplementary motor area; aSt, anterior striatum speech area; aT, anterior thalamus speech area; Av, avalanche; CMM, caudal medial mesopallium; CSt, caudal striatum; DM, dorsal medial midbrain nucleus; HVC, a letter-based name; L2, Field L2; dLMC, dorsal laryngeal motor cortex; vLMC, ventral laryngeal motor cortex; preLMC, premotor laryngeal motor cortex; OMC, oral motor cortex; MAN, magnocellular nucleus of the nidopallium; MO, mesopallium oval nucleus; NCM, nidopallium, caudal medial part; Nlf, nidopallium interfacial nucleus; NLC, nidopallium, lateral caudal; PAG, periaqueductal gray; RA, robust nucleus of the arcopallium; XIIts, 12th vocal motor nucleus, tracheosyringeal part. [Figure is updated and modified from (4)]



**Fig. 4. Evolutionary view of vocal learning in birds translated to spoken language in humans.** Shown is a summary of many labs. Each oval is a songbird song/vocal or respiratory nucleus. White circles, excitatory neurons; black circles, inhibitory neurons; plus sign, excitatory transmitter release; minus sign, inhibitory transmitter release. The RA-projection neurons of HVC fire in 10-ms synfire chains, where RA translates those sequences to control vocal motor and respiratory premotor neurons for the production of each 10 ms of sound through the syrinx (33). MAN and Area X inject variability and stereotypy, respectively, into both HVC and RA (4, 35). Predicted human brain regions and neuron cell types that correspond to the vocal learning circuits in songbirds are marked with a human outline. Abbreviations and color-coding are the same as Fig. 3. RA, retroambiguus; RVL, rostral nucleus of the ventral-lateral medulla.

Taken together, the most strongly supported differences in species that have imitative song and/or speech include: greater density of forebrain neurons that may accommodate additional song and speech brain circuits (hypothesis 1); an additional or enhanced forebrain vocal motor learning pathway (hypothesis 3); and a novel or enhanced forebrain-to-brainstem vocal motor connection (hypothesis 4). If stronger evidence were found for the other hypotheses, these would be additions to hypotheses 1, 3, and 4, rather than alternatives. The question that remains is how convergent evolution built similar brain pathways for a complex behavior.

### Motor theory of vocal learning origin

The finding that vocal learning and speech pathways in birds and humans are embedded in, or adjacent to, apparent motor learning pathways (17, 28) led to the motor theory of vocal learning origin, which posits that brain pathways for vocal learning and speech evolved independently from surrounding motor learning pathways found in all species and thus share a deep homology (28). The proposed mechanism is brain evolution by brain pathway duplication (Fig. 5) (28, 29). A motor learning pathway that already receives auditory input may be replicated multiple times during embryonic devel-

opment to innervate brainstem and spinal cord motor neurons for various muscle groups (Fig. 5B). In vocal learners, an extraneously duplicated pathway (Fig. 5C) could innervate brainstem vocal motor neurons and then be selected for a vocal learning phenotype (Fig. 5D). This vocal learning pathway is proposed to have been duplicated at least one other time, where the parrot inner core song system that is similar to songbirds and hummingbirds gave rise to the parrot shell song system unique to them (29), the human vLMC to the dLMC (17), and preLMC to Broca's area (29). Testing the pathway-duplication hypothesis may require experimental tools for lineage tracing of neural stem cell development.

### Convergent genetic changes for vocal learning brain circuits

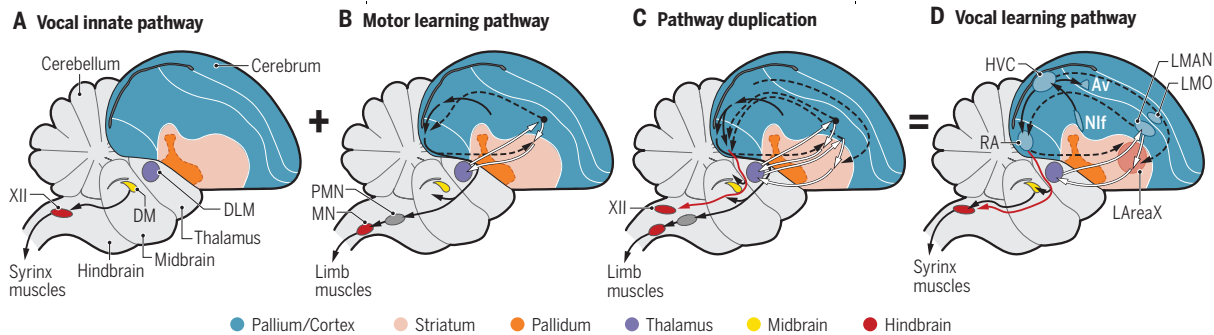
One prediction of the motor theory of vocal learning origin is that the vocal learning pathways should share molecular and functional similarities with adjacent motor learning pathways but diverge in certain neural connectivity genes, such as those that control a dense direct projection to brainstem vocal motor neurons. My colleagues and I tested this prediction by profiling the expression of thousands of genes in avian and primate brains (15). The gene

expression profiles supported the nuclear-to-layered hypothesis of avian and mammalian cortex relationships, where the avian arcopallium in which RA resides has cell types similar to mammalian motor cortex layer 5 neurons, and the nidopallium in which HVC resides has cell types similar to layers 2 or 3 (15) (Fig. 4). Gene expression profiles of avian song and human spoken-language brain regions resembled motor regions more than auditory regions (15) and diverged from the surrounding regions to become highly specialized. About 50 to 70 genes, many key to neural connectivity, per avian song and human spoken-language brain region showed convergent specialized expression.

Among down-regulated genes was the *SLIT1* axon guidance ligand (fig. S7, A to C). *SLIT1* interaction with its receptor *ROBO1* prevents axon connections from forming. We proposed that down-regulation of *SLIT1* in songbird RA and human LMC layer 5 neurons may produce a permissive environment for their axons to form dense direct projections to brainstem vocal motor neurons with high levels of *ROBO1* (fig. S7D). This concept is supported by the *PlexinA1* down-regulation in cortical layer 5 and its increased connections to forelimb motor neurons mentioned earlier (21). Mutations in the *ROBO1* locus are associated with dyslexia and speech sound disorders (30). Mutations in *FOXP2*, a transcription factor that directly regulates *SLIT1*, cause a phoneme sequence speech deficit in humans and a similar but more rudimentary syllable sequence deficit in mice, associated with less localized LMC layer 5 neurons (31). In humans, partial gene duplications of the *SLIT-ROBO* guanosine triphosphatase 2 (*SRGAP2*) gene (32) encode proteins that act as competitive inhibitors of full-length *SRGAP2*, keeping synapses at a higher density and more plastic into adulthood (fig. S2C). All these findings suggest that not only convergent changes in the same genes between species but also on different genes in the same genetic pathway within a species are associated with the evolution of vocal learning and spoken language.

### Vocal learning continuum hypothesis in three anatomical stages

I suggest that the common ancestor of vertebrates had a brainstem pathway for production of innate vocalizations with limited vocal plasticity, such as the Lombard effect, where animals increase sound production volume or pitch in noisy environments (fig. S8). In some species, the forebrain motor learning pathway then duplicated and formed a vocal motor learning pathway with weak direct projections to the brainstem vocal motor neurons. Thereafter, this forebrain vocal motor learning pathway expanded in neuron numbers causing greater density of neurons in the forebrain, moved outside of the motor learning pathway, and gained dense direct projections to brainstem vocal motor



**Fig. 5. Brain pathway duplication origin of vocal learning motor pathway.**

(A) Innate vocal brainstem pathway found in all vertebrate species that vocalize, shown here for birds. (B) Motor learning pathway found in all vertebrate species. (C) Proposed additional forebrain motor learning pathway duplication that connects to the innate brainstem vocal pathway. (D) Resultant vocal learning

pathway in songbirds, which has similarities to the surrounding motor learning pathway in (B) connected to the innate vocal pathway in (A). Abbreviations and arrow color-coding are the same as Fig. 3. MN, motor neuron; PMN, premotor neurons in reticular formation; LMAN, lateral MAN; LMO, lateral MO. [Hypothesis based on (28, 29)]

neurons. Finally, the vocal learning pathway then duplicated one or more times and took on additional specialized gene regulation and connections, resulting in the advanced vocal learning pathways we find in parrots and in humans.

### Brain mechanisms of vocal learning and spoken language

The evolution-based findings allow us to make predictive translations between species. For example, the HVC-to-RA projection neurons of songbirds fire sparsely in a synfire chain with 10-ms time resolution thought to sequence sounds within and among syllables (33) (Fig. 4). RA in turn translates the forebrain signals to the brainstem vocal and respiratory neurons to dictate acoustic structure of syllables. The HVC-to-Area X projection neurons send a corollary efference copy of the sequence into the striatal Area X vocal nucleus of the anterior forebrain pathway (Fig. 4) (34). If the human spoken-language pathway functions in a similar manner, then I predict that: (i) human LMC layer 2-3 neurons fire in a synfire chain (like HVC) onto the LMC layer 5 neurons (like RA) to control a defined millisecond time resolution for producing learned phoneme and word sequences; and (ii) some LMC neurons send an efference copy to the anterior speech striatum (Figs. 3B and 4). During singing, songbird MAN neurons inject variable neural activity into the vocal motor pathway (to RA and HVC), and Area X in the striatum modulates or constrains that variability (Fig. 4) (35). Similarly, I predict that human layer 2-3 motor neurons of premotor LMC or Broca's may innervate both dLMC and vLMC to inject acoustic variability in speech production, and the anterior striatum modulates that variability (Figs. 3B and 4).

Predictions from songbirds extend to human molecular and neurophysiological mechanisms. In songbirds, cells of the vocal learning pathway show diverse activity-dependent patterns of up- or down-regulation of hundreds to thousands of genes in several temporal waves after singing that define function of different cell types (36).

Some of the genes are specialized to the vocal learning circuit (37). I predict that humans may also have song- and speech-driven gene regulation in the spoken-language brain pathway, in a cell type-specific manner (Figs. 3B and 4). Neurophysiology experiments show that auditory responses in the vocal learning pathway are suppressed when songbirds sing (38). In humans as well as nonhuman primates (marmoset), there is also suppression in the auditory cortex with vocalization (16, 39). In mice, auditory information enters the motor cortex (Fig. S5, D to F) and, like in the songbird vocal learning pathways, is gated off in motor regions when the mice move (40). This suggests that the suppression of auditory input into the vocal pathway during vocalization in vocal learners is an ancestral trait inherited from the adjacent motor pathway.

Certain mechanisms will not be shared because of differences in avian and mammalian cortical organization, cell types, and peripheral musculature. Despite these differences, underlying principles, as outlined here, do translate across different systems. Although song and speech pathways in vocal learning birds and humans are specialized, the majority of genes, neural connections, and physiology are similar to their adjacent brain pathways.

### REFERENCES AND NOTES

- W. T. Fitch, *Psychon. Bull. Rev.* **24**, 3–33 (2017).
- C. I. Petkov, E. D. Jarvis, *Front. Evol. Neurosci.* **4**, 12 (2012).
- M. Perlman, N. Clark, *Anim. Cogn.* **18**, 1165–1179 (2015).
- E. D. Jarvis, *Ann. N. Y. Acad. Sci.* **1016**, 749–777 (2004).
- C. N. Templeton, E. Greene, K. Davis, *Science* **308**, 1934–1937 (2005).
- G. Arriaga, E. D. Jarvis, *Brain Lang.* **124**, 96–116 (2013).
- A. R. Lameira, *Neurosci. Biobehav. Rev.* **83**, 429–439 (2017).
- F. Patterson, E. Linden, *The Education of Koko* (Henry Holt & Company, 1981).
- J. L. Russell, J. M. McIntyre, W. D. Hopkins, J. P. Tagliavola, *Brain Lang.* **127**, 520–525 (2013).
- K. E. Bouchard, N. Mesgarani, K. Johnson, E. F. Chang, *Nature* **495**, 327–332 (2013).
- T. Riede, F. Goller, *Brain Lang.* **115**, 69–80 (2010).
- S. Herculano-Houzel, *Front. Hum. Neurosci.* **3**, 31 (2009).
- W. T. Fitch, *Annu. Rev. Linguist.* **4**, 255–279 (2018).
- S. M. Garcia et al., *Curr. Biol.* **27**, 2677–2683.e3 (2017).
- A. R. Penning et al., *Science* **346**, 1256846 (2014).

- B. K. Dichter, J. D. Breshears, M. K. Leonard, E. F. Chang, *Cell* **174**, 21–31.e9 (2018).
- M. Belyk, S. Brown, *Neurosci. Biobehav. Rev.* **77**, 177–193 (2017).
- K. Simonyan, *Curr. Opin. Neurobiol.* **28**, 15–21 (2014).
- C. M. Cerkevich, P. L. Strick, poster presented at the Society for Neuroscience Annual Meeting, Chicago, IL, 17 to 21 October 2015.
- W. C. Liu, K. Wada, E. D. Jarvis, F. Nottebohm, *Nat. Commun.* **4**, 2082 (2013).
- Z. Gu et al., *Science* **357**, 400–404 (2017).
- A. D. Friederici, N. Chomsky, R. C. Berwick, A. Moro, J. J. Bolhuis, *Nat. Hum. Behav.* **1**, 713–722 (2017).
- J. D. Breshears, D. G. Southwell, E. F. Chang, *Neurosurgery* **85**, E496–E501 (2019).
- A. Basilakos, K. G. Smith, P. Fillmore, J. Fridriksson, E. Fedorenko, *Cereb. Cortex* **28**, 1816–1830 (2018).
- B. Alderson-Day, C. Fernyhough, *Psychol. Bull.* **141**, 931–965 (2015).
- K. Emmorey et al., *Neuroimage* **17**, 812–824 (2002).
- B. K. Young, G. B. Mindlin, E. Arnedo, F. Goller, *PeerJ* **5**, e4052 (2017).
- G. Feenders et al., *PLOS ONE* **3**, e1768 (2008).
- M. Chakraborty, E. D. Jarvis, *Philos. Trans. R. Soc. London Ser. B* **370**, 20150056 (2015).
- K. Hannula-Jouppi et al., *PLOS Genet.* **1**, e50 (2005).
- J. Chabout et al., *Front. Behav. Neurosci.* **10**, 197 (2016).
- C. Charrier et al., *Cell* **149**, 923–935 (2012).
- R. H. R. Hahnloser, A. A. Kozhevnikov, M. S. Fee, *Nature* **419**, 65–70 (2002).
- T. W. Troyer, A. J. Doupe, *J. Neurophysiol.* **84**, 1204–1223 (2000).
- M. S. Fee, J. H. Goldberg, *Neuroscience* **198**, 152–170 (2011).
- O. Whitney et al., *Science* **346**, 1256780 (2014).
- H. Horita et al., *PLOS ONE* **7**, e42173 (2012).
- K. Hamaguchi, K. A. Tschida, I. Yoon, B. R. Donald, R. Mooney, *eLife* **3**, e01833 (2014).
- S. J. Eliades, X. Wang, *Hear. Res.* **348**, 98–111 (2017).
- D. M. Schneider, R. Mooney, *Curr. Opin. Neurobiol.* **33**, 78–84 (2015).

### ACKNOWLEDGMENTS

I thank members of the Jarvis lab for critical discussions and comments, especially C. Theofanopoulou, L. Shalmiyev, C. Vargas, and G. Gedman for critical reading and discussions of the manuscript. I thank E. Chang, M. Long, and K. Emmorey for critical discussions on their work on brain electrophysiology and imaging during spoken-language, sign-language, and other tasks in humans. I thank other members of the animal and human communication communities and the brain evolution community for valuable discussions over the years that have helped formulate the hypotheses proposed. I apologize to those whose work I did not cite due to the need to keep citations to a limit. Most of the additional relevant studies can be found in the papers cited. **Funding:** The author's efforts are supported by funds from the Howard Hughes Medical Institute, a NIH Director's Transformative Research Award, and The Rockefeller University. **Competing interests:** The author declares no competing interests.

### SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6461/50/suppl/DC1  
Supplementary Text  
Figs. S1 to S8  
References (41–60)

10.1126/science.aax0287

## Evolution of vocal learning and spoken language

Erich D. Jarvis

*Science* **366** (6461), 50-54.  
DOI: 10.1126/science.aax0287

<b>ARTICLE TOOLS</b>	<a href="http://science.sciencemag.org/content/366/6461/50">http://science.sciencemag.org/content/366/6461/50</a>
<b>SUPPLEMENTARY MATERIALS</b>	<a href="http://science.sciencemag.org/content/suppl/2019/10/02/366.6461.50.DC1">http://science.sciencemag.org/content/suppl/2019/10/02/366.6461.50.DC1</a>
<b>RELATED CONTENT</b>	<a href="http://science.sciencemag.org/content/sci/366/6461/48.full">http://science.sciencemag.org/content/sci/366/6461/48.full</a> <a href="http://science.sciencemag.org/content/sci/366/6461/55.full">http://science.sciencemag.org/content/sci/366/6461/55.full</a> <a href="http://science.sciencemag.org/content/sci/366/6461/58.full">http://science.sciencemag.org/content/sci/366/6461/58.full</a> <a href="http://science.sciencemag.org/content/sci/366/6461/62.full">http://science.sciencemag.org/content/sci/366/6461/62.full</a> <a href="http://science.sciencemag.org/content/sci/366/6461/13.full">http://science.sciencemag.org/content/sci/366/6461/13.full</a> <a href="http://science.sciencemag.org/content/sci/366/6461/33.full">http://science.sciencemag.org/content/sci/366/6461/33.full</a> <a href="http://science.sciencemag.org/content/sci/366/6461/83.full">http://science.sciencemag.org/content/sci/366/6461/83.full</a>
<b>REFERENCES</b>	This article cites 58 articles, 5 of which you can access for free <a href="http://science.sciencemag.org/content/366/6461/50#BIBL">http://science.sciencemag.org/content/366/6461/50#BIBL</a>
<b>PERMISSIONS</b>	<a href="http://www.sciencemag.org/help/reprints-and-permissions">http://www.sciencemag.org/help/reprints-and-permissions</a>

Use of this article is subject to the [Terms of Service](#)

---

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works